Title page
&
Index

DUKE MATHEMATICAL JOURNAL

EDITED BY

ARTHUR BYRON COBLE   DAVID VERNON WIDDER

JOSEPH MILLER THOMAS
Managing Editor

WITH THE COÖPERATION OF

VOLUME 1
1935

DUKE UNIVERSITY PRESS
DURHAM, N. C.

CONTENTS

VOLUME 1, 1935

CHAIN-DEFORMATIONS IN TOPOLOGY

By S. Lefschetz

In topology one has repeated occasion to consider homotopic deformations of chains. They give rise to a basic boundary relation[1] between the extreme positions of a chain c_p in the homotopy and what might be termed the loci of c_p and of its boundary $F(c_p)$. All the consequences of the homotopy that concern algebraic topology (i.e., boundary relations and the associated homologies) may be derived from the fundamental relation. It seems natural therefore to call *chain-deformation* any scheme wherein two p-chains c_p, c_p' and two other chains that are to take the part of the loci mentioned above, satisfy a boundary relation formally identical with the fundamental relation of homotopy.[2] This notion has already been exploited in a recent paper.[3] We return to it here, first to develop it more fully and then to apply it to the study of the sets that are obtained whenever, in the definition of locally connected sets, singular cells and spheres are replaced by chains. These new sets may be described as locally connected in the sense of homology, and their types correspond substantially to the locally connected types that we have recently investigated.[4] The passage from the first class to the second corresponds also to a substitution of chain-deformation for homotopy.

One of the important results of L2 was the identification of certain locally

Received February 12, 1935.

[1] Given for the first time in our Colloquium Lectures, *Topology*, New York, 1930, p. 78.

[2] While chain-deformations have most of the properties that their name suggests, they are essentially different from homotopy. This is clearly seen by noting the different effect in the very simple case of the circuits on an orientable surface of genus $p \geqq 2$. Homotopy leads, in this case, to the non-commutative Poincaré group, chain-deformation to the much simpler abelian group with $2p$ free generators.

[3] S. Lefschetz, *On generalized manifolds* (= L1 in the sequel), American Journal of Mathematics, vol. 55 (1933), pp. 475–499.

[4] S. Lefschetz, *On locally connected and related sets* (= L2 in the sequel), Annals of Mathematics, vol. 35 (1934), pp. 118–139. We call attention to the following errata: p. 119, line 23, replace LC by LC^∞; p. 126, suppress line 3 from bottom; in line 4 from bottom, suppress "convex"; in line 5 from bottom, replace "convex sets of $\mathfrak{H}$" by "spheres"; p.127, line 13, replace K^* by $\mathfrak{K}^*$.

Local connectedness in the sense of homology was introduced by P. S. Alexandroff in his paper: *Untersuchungen über Gestalt und Lage abgeschlossener Mengen beliebiger Dimension*, Annals of Mathematics, vol. 30 (1929), pp. 101–187. See also his recent paper: *On local properties of closed sets*, Annals of Mathematics, vol. 36 (1935), pp. 1–35, §3. The same property for euclidean domains plays a central part in R. L. Wilder's recent work. See in particular his last paper: *Generalized closed manifolds in n-space*, Annals of Mathematics, vol. 35 (1934), pp. 876–903.

connected sets with the absolute neighborhood retracts or absolute retracts in the sense of Borsuk. (See in this connection footnote 13.) A similar identification is possible here with generalized retracts in the sense of chain-deformations. Rather than to press this analogy, we preferred to investigate the mutual relations between the two kinds of local connectedness as well as with Borsuk's locally contractible sets, but there remains still much to be done along that line.

One of the most useful notions introduced in L2 was that of the semi-singular complex (singular complex with only part of the expected cells present). It is extended here to aggregates of chains related like the oriented cells of a complex and is found no less useful in the present investigation.

In the endeavor to free our results, as far as possible, from any specific choice of chains, we have presented the theory of chains in axiomatic form at the beginning of the paper. This has the additional advantage of making the paper less dependent upon our previous writings.

Our general notations are those of *Topology*, with the abbreviations of L1 and L2: LC, and NR stand for "locally connected" and "neighborhood retract." In addition we shall write HLC for "LC in the sense of homology", and similarly for HNR. Our two other abbreviations are c.s.v.t. for "continuous single-valued transformation", f.c.o.s. for "finite covering by open sets". The reader will have no difficulty in getting accustomed to these alphabetical notations whose advantages, after all, need not be reserved for the political domain.

§1. The chains of a topological space

1. There are various ways of extending to a topological space $\mathfrak{R}$ the basic properties of the chains of a geometric complex, their cycles, their boundary relations and the like. Regardless of the procedure adopted certain properties are preserved. As it is with these common properties that we are chiefly concerned, we shall recall them briefly and state them as postulates for the chains of $\mathfrak{R}$:[5]

I. *There exists for every* $p = 0, 1, \cdots$, *a set of topological invariants of* $\mathfrak{R}$, *its p-chains* c_p, *and their collection* $\{c_p\}$ *constitutes a free additive abelian group.*

II. *There exists an operation F defined topologically for all the chains and such that* $F\{c_p\}$ *is a homomorphism of* $\{c_p\}$ *into* $\{c_{p-1}\}$, *and of* $\{c_0\}$ *into the identity.*

F is the *boundary-operator*, Fc_p or $F(c_p)$ is the *boundary* of c_p, their mutual relation being indicated by a *boundary relation*,

$$c_p \rightarrow F(c_p), \tag{1.1}$$

[5] These properties are fully developed in *Topology*, Chapters I, II. For a more systematic exploitation of the abstract viewpoint see A. W. Tucker, *An abstract approach to manifolds*, Annals of Mathematics, vol. 34 (1933), pp. 191–243, where further references, notably to the papers of W. Mayer, will be found. It is to be noted that whereas they consider only abstract complexes and chains, we have always tied them up with definite point-sets. It might be advisable to use different terms, such as abstract complex or chain, geometric complex or chain, for the two concepts.

while to express merely that c_p is a boundary we write a *homology*,

$$c_p \sim 0. \tag{1.2}$$

The chains c_p such that $Fc_p = 0$ are called *p-cycles*, and generically denoted by γ_p. In particular, every c_0 is a γ_0. From II follows that $\{\gamma_p\}$ is a subgroup of $\{c_p\}$.

III. $FF = 0$.

This means that boundaries are cycles. Moreover if β_p is a generic boundary, from II follows again that $\{\beta_p\}$ is a subgroup of $\{\gamma_p\}$. The difference-group (factor-group of the customary terminology) $\{\gamma_p\} - \{\beta_p\}$ is the p^{th} *homology group* of $\mathfrak{R}$. When it is a free group the number $R_p(\mathfrak{R})$ of its independent generators is called the p^{th} *Betti-number* of $\mathfrak{R}$.

It is to be kept in mind that in the present paper, *all homologies imply bounding*. This is the reason why we use the symbol $\sim$ and not $=$, for example as in *Topology*, Chapter VII, or L1, for $\gamma_p = 0$ merely meant that γ_p was a finite or infinite sum of bounding cycles, or neglected chains, without being itself strictly in one or the other category. From the group viewpoint, our earlier procedure corresponds to topologizing the groups, replacing $\{\beta_p\}$ by its closure, say $\{\beta_p'\}$, and taking as p^{th} homology group the difference-group $\{\gamma_p\} - \{\beta_p'\}$.

IV. *There exists a numerical topological invariant linear function of zero-chains, the Kronecker-index* (c_0), *and* $(c_0) = 0$ *when* $c_0 \sim 0$.

In the case of complexes (c_0) is the number of points of c_0 each counted with its coefficient in the expression of the chain.

V. *With every* c_p *there is associated a unique closed subset* $|c_p|$ *of* $\mathfrak{R}$ *such that:*

(a) $|0| = 0$; (b) $|c_p + c_p'| \subset |c_p| + |c_p'|$,

(c) $|F(c_p)| \subset |c_p|$; (d) $\dim |c_p| \geqq p$ when $c_p \neq 0$.

If A is any closed subset of $\mathfrak{R}$ we say that $c_p \subset A$ whenever $|c_p| \subset A$. Taken together with V, this enables us to define the boundary relations, homologies and homology-groups mod A, or *relative* relations, by contrast with the previous type called *absolute*.

From V(d) follows that *when* $\dim \mathfrak{R} = n$ *is finite every* $c_p = 0$ *for* $p > n$. In particular there are no homology groups, absolute or relative, for dimensions $> n$.

2. There remains one more property, but it is most conveniently expressed in terms of the very useful notion of *quasi-complex*. A quasi-complex $\mathfrak{K}$ is a collection of chains of $\mathfrak{R}$ such that: (a) its p-chains form a subgroup of $\{c_p\}$; (b) $F\,\mathfrak{K} \subset \mathfrak{K}$: when $c_p \subset \mathfrak{K}$ likewise $Fc_p \subset \mathfrak{K}$. The quasi-complex is *finite* whenever the dimension of its chains is bounded, and in addition for every p there is a finite base for its p-chains whose elements are independent. In that case we frequently reserve the name "chain of $\mathfrak{K}$" for the chains of the bases, and call *subchains* of $\mathfrak{K}$ the other chains of the quasi-complex.

If c_q^i is any base chain of a finite $\mathfrak{K}$, the chains c_{q-1}^j entering in the composition of $F(c_q^i)$, those entering in the composition of $F(c_{q-1}^j)$, etc., are called the *boundary-chains* of c_q^i. The sum $|c_q^i| + \Sigma\,|c_{q-1}^j| + \cdots$ has a least upper bound called the *mesh* of $\mathfrak{K}$.

Our last axiom may now be stated:

VI. (*Subdivision axiom.*) *Every chain of $\mathfrak{R}$ is a subchain of a finite quasi-complex $\mathfrak{K}$ whose mesh is arbitrarily small.*

The complex $\mathfrak{K}$ is called an *elementary decomposition* of the chain, an elementary ϵ-decomposition when its mesh $< \epsilon$.

3. The following are noteworthy examples of systems of chains for which all the axioms hold:

(α) $\mathfrak{R}$ is a topological space and its chains are the singular chains on $\mathfrak{R}$ in the sense of *Topology*, Chapter II,[6] that is to say, the linear forms in the singular cells on $\mathfrak{R}$ with coefficients members of an additive abelian group $\mathfrak{M}$. All the axioms except the last are readily verified, and the last is verified also provided that we agree to identify a singular c_p with all its subdivisions and call *chain* the class thus obtained. Otherwise we merely have as a theorem that all the members of the class are equivalent regarding boundary relations and homologies (*Topology*, p. 88). It is understood of course that each cell of a subdivision of c_p is to be oriented concordantly with the carrying cell of c_p.

A noteworthy case is that where $\mathfrak{R}$ is a simplicial complex K. It is then shown that the homology groups derived from the chains made up of the oriented simplexes of K, or *combinatorial* homology groups, are isomorphic with those defined above and hence the former are topological invariants.

(β) $\mathfrak{R}$ is a compact metric space and the chains are the *projection-chains* of L1. They are certain specific subchains, taken here with coefficients in a field $\mathfrak{M}$, of a fundamental infinite complex K which consists of the skeleta Φ^i of f.c.o.s. together with their joining cells (L1, p. 470–472).[7] The configuration $\bar{K} = \mathfrak{R} + K$ may be identified for convenience with its topological image on the Hilbert parallelotope $\mathfrak{H}$ (see loc. cit. for details). The projection-chains C_{p+1} of K determine the chains c_p of $\mathfrak{R}$, with $| c_p | = \overline{| C_{p+1} |} \cdot \mathfrak{R}$ (L1, p. 477). For $p = 0$ we define the Kronecker-index by $(c_0) = (F(C_1))$ which is equivalent to the definition of L1, p. 489.

We shall call chains of type (α) *singular* and chains of type (β) *regular*. Unless otherwise specified regular chains shall be the type usually considered in the sequel.

The only non-compact metric spaces that interest us are the separable locally compact spaces. For these we may consider the so-called *finite* cycles, or cycles on self-compact subsets of the space, and they are the only kind needed later.

Our definition of regular chains is not intrinsic in that it depends upon a specific fundamental complex K. It may be made intrinsic as follows. Suppose that we pass to a new fundamental complex K'. By means of the deformation theorem of *Topology*, p. 328 we

[6] See also S. Lefschetz, *On singular chains and cycles*, Bulletin American Mathematical Society, vol. 39 (1933), pp. 124–129.

[7] In the construction of K (loc. cit., beginning of No. 2) the finiteness of dim $\mathfrak{R}$ plays no rôle. All that matters is that the number n is the least order of any ϵ —f.c.o.s.

first reduce a projection-chain C_{p+1} of K defining c_p of $\mathfrak{R}$ to a chain C'_{p+1} of K', then by Theorems IV, V of L1 (p. 484), we reduce C'_{p+1} to a projection-chain C''_{p+1} of K'. The construction, which depends upon a choice of bases as in L1, No. 16, is made unique by adopting a fixed numbering of all the cells of K'. It is the class of all projection-chains C''_{p+1} thus obtained which is c_p under the new definition. It is not clear, however, that $|c_p|$ is independent of K. Let us assume that this is not necessarily so and let A, A' be the sets as determined respectively by K and K'. We have immediately $A \subset A'$. If we return from K' back to K we show readily that we reduce C''_{p+1} to a projection-chain on a fundamental complex for A necessarily identical with C_{p+1}. Hence $A' \subset A$ and therefore $A' = A$.

It may be observed that projection-chains are imposed by the subdivision axiom. They are in fact the only chains of K which we are able to subdivide into elements of the same nature, notably as regards the subdivision of the elements C_2 into elements with the proper C_1 boundaries. On the other hand their uniqueness is only proved when $\mathfrak{M}$ is a field, so that the restriction imposed upon $\mathfrak{M}$ is largely due, in the last analysis, to the subdivision axiom.

Besides the two classes (α), (β) others may well be introduced. We mention notably: (γ) the Vietoris-chains for compact metric spaces;[8] (δ) the chains recently defined by Čech[9] for spaces even more general than topological spaces. In the case (γ) there is an associated group $\mathfrak{M}$ as general as in (α), while in (δ) $\mathfrak{M}$ is again restricted to a field. In fact, when $\mathfrak{M}$ is a field the homology theories yielded by (β), (γ), (δ) are the same. Furthermore when $\mathfrak{M}$ is a field, and $\mathfrak{R}$ is LC^{∞} in the sense of L2 the four types coalesce (*Topology*, p. 333).

All but the last of our axiomatic properties are readily verified for the four types mentioned. Regarding the last, subdivision, for the singular type it is a consequence of the definition and for the regular type it is proved as in L1, p. 490. Since they are the only two considered in the present paper this will suffice for our purpose.

§2. Chain-deformations

4. As a matter of convenience we shall assume henceforth that the basic space $\mathfrak{R}$ is separable, metric, locally compact, and later even, a good part of the time, that it is compact. Until further notice the chains are finite chains in the sense of No. 3 and, unless otherwise stated, merely subjected to our basic axioms.

In the applications, even when dealing with a single chain, one is constrained to deform a whole quasi-complex. For example to have the analogue of ϵ-homotopy, we must break up the chains into the elements of a $\mathfrak{K}$ of mesh ϵ which is then to undergo the deformation. Owing to this it is best to define at the outset a *chain-deformation* ϑ of a quasi-complex $\mathfrak{K} = \{c_q\}$, into another $\mathfrak{K}' = \{c'_q\}$. We understand then by such a chain-deformation, (merely "deformation" when there is no ambiguity), a homomorphism of the chain-groups of $\mathfrak{K}$ into those of the same dimension for $\mathfrak{K}'$, which commutes with F and which is associated with a linear operator $\mathfrak{D}$ on the chains of $\mathfrak{K}$ having the property that

[8] L. Vietoris, *Über den höheren Zusammenhang kompakter Räume und eine Klasse von zusammenhangstreuen Abbildungen*, Math. Annalen, vol. 97 (1927), pp. 454–472.

[9] E. Čech, *Théorie générale de l'homologie dans les espaces abstraits*, Fundamenta Mathematicae, vol. 19 (1932), pp. 149–183.

whatever c_q of $\mathfrak{K}$, $\mathfrak{D}c_q$ is a uniquely defined $(q + 1)$-chain of $\mathfrak{R}$, called the *deformation-chain* of c_q, such that

$$F\,\mathfrak{D}\,c_q = \vartheta c_q - c_q - \mathfrak{D}\,F\,c_q. \tag{4.1}$$

Written as relations between operators on $\mathfrak{K}$ we have then

$$\vartheta\,F = F\,\vartheta, \tag{4.2}$$

$$F\,\mathfrak{D} + \mathfrak{D}\,F = \vartheta - 1. \tag{4.3}$$

$\mathfrak{D}$ may also be considered as inducing for each q a homomorphism of the group of the q-chains of $\mathfrak{K}$ into the group of the $(q + 1)$-chains of $\mathfrak{R}$ satisfying (4.3).

The preceding properties, and in particular (4.1), written also

$$\mathfrak{D}\,c_q \to \vartheta c_q - c_q - \mathfrak{D}\,F(c_q), \tag{4.4}$$

are those satisfied by chains under a homotopy (*Topology*, p. 78) which we are thus merely taking as postulates for chain-deformation.

5. A cycle γ_p, a chain c_p + its boundary $F(c_p)$, form special quasi-complexes. For the former ϑ merely demands that

$$F\,\mathfrak{D}\,\gamma_p = \vartheta\gamma_p - \gamma_p. \tag{5.1}$$

Since a boundary is a cycle, $\vartheta\gamma_p - \gamma_p$ is a cycle, and hence $\vartheta\gamma_p$ is one also. Therefore *the chain-deform of a cycle is a cycle.* Regarding c_p, ϑ demands two deformation-chains $\mathfrak{D}\,c_p$, $\mathfrak{D}\,Fc_p$ such that

$$F\,\mathfrak{D}\,c_p = \vartheta c_p - c_p - \mathfrak{D}\,Fc_p, \tag{5.2}$$

$$F\,\mathfrak{D}\,Fc_p = F\vartheta c_p - Fc_p. \tag{5.3}$$

Of these relations the second follows from the first, since it merely expresses the fact that the boundary of the right-hand side in (5.2) is a cycle. Therefore for c_p also a chain-deformation is specified by the single relation (5.2). It is clear that in the sense just considered a chain-deformation of a quasi-complex $\mathfrak{K}$ induces a chain-deformation of its individual chains.

6. Whenever $\mathfrak{K}$, $\mathfrak{K}'$ and the $\mathfrak{D}$-chains are on a given set A we say that the chain-deformation is *over* A. The chains in any given set B form a quasi-complex, and its *chain-deformations* are called *chain-deformations* of B. Thus if ϑ deforms all the chains of B into chains of a set B over A, we say that B is chain-deformed into B'.

Whenever $\mathfrak{K}$, $\mathfrak{K}'$ are of mesh $< \epsilon$ and all the chains $\mathfrak{D}$ involved are of diameter $< \epsilon$, we say that ϑ is an ϵ chain-deformation of $\mathfrak{K}$ or of any subchain of $\mathfrak{K}$. An ϵ chain-deformation of c_p consists in imposing upon it an elementary ϵ-decomposition reducing it to a $\mathfrak{K}$ of mesh $< \epsilon$, and then ϵ chain-deforming $\mathfrak{K}$.

7. Let ϑ be as before and let ϑ' deform $\mathfrak{K}'$ into $\mathfrak{K}''$ with $\mathfrak{D}'c_q'$ as the deformation-chains. We have now a transformation $\vartheta'' = \vartheta'\vartheta$ of $\mathfrak{K}$ into $\mathfrak{K}''$ and we shall show that it is a chain-deformation. Clearly ϑ'' is a homomorphism commuting with F so (4.3) alone must be verified. We have

$$\mathfrak{D}'F + F\mathfrak{D}' = \vartheta' - 1 \tag{7.1}$$

operating on $\mathfrak{K}'$, but considered as operating on $\mathfrak{K}$ this must be written

$$\mathfrak{D}'F\vartheta + F\mathfrak{D}'\vartheta = \vartheta'\vartheta - \vartheta = \vartheta'' - \vartheta. \tag{7.2}$$

Adding (7.2) and (4.3), we have

$$(\mathfrak{D} + \mathfrak{D}'\vartheta)F + F(\mathfrak{D} + \mathfrak{D}'\vartheta) = \vartheta'' - 1. \tag{7.3}$$

Hence if we introduce the linear chain-operator $\mathfrak{D}'' = \mathfrak{D} + \mathfrak{D}'\vartheta$, we have

$$\mathfrak{D}''F + F\mathfrak{D}'' = \vartheta'' - 1, \tag{7.4}$$

so that (4.3) holds, with $\mathfrak{D}''$ as the $\mathfrak{D}$ operation. From its definition $\mathfrak{D}''$ is linear, hence $\mathfrak{D}''$ is a deformation-chain.

Clearly $\vartheta = 1$ is a chain-deformation corresponding to $\mathfrak{D} = 0$. Regarding the inverse ϑ^{-1} it can only be defined, if at all, when ϑ is one-one. In particular, this holds when $\mathfrak{K}$ consists of a single chain plus its boundary. When ϑ is one-one $c_q' = \vartheta c_q$ determines uniquely c_q and hence $\mathfrak{D}\, c_q$. If we set

$$\mathfrak{D}'c_q' = \mathfrak{D}'\vartheta\, c_q = -\mathfrak{D}\, c_q$$

we have

$$\mathfrak{D}'\vartheta Fc_q = \mathfrak{D}'F\vartheta c_q = -\mathfrak{D}Fc_q,$$

and therefore

$$F\mathfrak{D}'c_q' = c_q - c_q' - \mathfrak{D}\, F(c_q').$$

This shows that the passage from $\mathfrak{K}'$ to $\mathfrak{K}$ is a chain-deformation with $-\mathfrak{D}c_q$ as its deformation-chain. Thus in the case in question ϑ^{-1} is a chain-deformation.

If we call two q-chains of $\mathfrak{R}$ *equivalent* whenever one of them can be chain-deformed into the other, the results of the two preceding paragraphs show that this equivalence is *transitive*, *reflexive*, and *symmetric*.

8. **Homotopy and chain-deformation.** Let a set A undergo a homotopy T into a set A' over $\mathfrak{R}$ and let B be its locus throughout the deformation. If c_p is a chain of A the homotopy will determine a chain $c_p' = Tc_p$. One suspects intuitively that, as is obvious for singular chains, c_p' is a chain-deform of c_p over B. We shall in fact prove:

THEOREM I. *When the chains adopted are regular or singular a homotopic deformation of a compact*[10] *set induces a chain-deformation of all its chains.*

[10] We mean here that the set is compact as a space, or, as it is sometimes called, self-compact.

For singular chains the theorem is practically proved in *Topology*, p. 79, so that we may limit our treatment to regular chains. This being the case, we may clearly replace A, B, A' by $| c_p |$, its locus, and $T \cdot | c_p |$, and hence assume A, A' closed and the locus the whole space, that is $B = \mathfrak{R}$. Under the circumstances let L be a fundamental complex for A which is a subcomplex of K (L1, p. 474) and let the whole configuration be assumed immersed in the Hilbert parallelotope $\mathfrak{H}$ (No. 3). Let $P^1, P^2, \cdots$ be the vertices of K, which we take to be points of A and consider the product $L \times \lambda = L^*$, where λ is a unit-segment. It is an infinite convex complex, which we subdivide in such a manner that if E is any cell of L, the cell $E \times \lambda$ of L^* becomes a sum of cells $E \times \lambda'$, where λ' is an interval of λ. That is, every "prismatic" cell of L^* is subdivided by sections "parallel" to its bases. This is carried out in such manner that if the cells of K, numbered in some order, are $E^1, E^2, \cdots$, then $E^k \times \lambda$ is subdivided into cells equally spaced (i.e. with intervals λ' of equal length) and whose number $> 2^k$. Finally we subdivide in any manner L^* simplicially without introducing new vertices, which may always be done since its cells are all convex. We continue to call L^* the ultimate complex. On L^* we have a representative C_{p+1} of c_p, and also a chain C'_{p+1} obtained by translation from C_{p+1} and deformation-chains $\mathfrak{D}\, C_{p+1}$, $\mathfrak{D} F C_{p+1}$ such that

$$\mathfrak{D}\, C_{p+1} \rightarrow C'_{p+1} - C_{p+1} - \mathfrak{D}\, F(C_{p+1}).$$

We shall now transform L^* barycentrically as follows. Let us suppose that the deformation T of A into A' depends upon a parameter t varying from 0 to 1. On the cell $P^k \times \lambda$ there will be a certain number of vertices $P^{k0}, \cdots, P^{kr}$ of L^*. Mark now on the path of P^k as t varies, the positions of P^k corresponding to the values i/r, $(i = 0, 1, \cdots, r)$, and let them be $P^k = Q^{k0}, \cdots, Q^{kr}$. We map P^{ki} into Q^{ki}, thus obtaining a unique image for every vertex of L^*, and for every simplex with certain vertices P^{ki} we insert the corresponding simplex with vertices Q^{ki}. The various chains on L^* are thus mapped into chains defining chains, of one dimension less on $\mathfrak{R}$, in the same sense as those on K (*Topology*, p. 324). Furthermore by a deformation in $\mathfrak{H}$, (*Topology*, p. 332) together with Theorems IV, V of L1, all leaving C_{p+1} invariant, we may reduce them to projection-chains of K. By the very definition of the different chains, $\mathfrak{D}\, C_{p+1}$ and C'_{p+1} go into chains of K which define a $\mathfrak{D}\, c_p$ and c'_p satisfying (4.1), with $| \mathfrak{D}\, c_p | \subset \mathfrak{R}$, $| c'_p | \subset A'$. This proves the theorem.

§3. Retraction properties

9. Henceforth we restrict the chains definitely to the *regular* or *singular* types, and until further notice, we also assume that the space $\mathfrak{R}$ is *compact, metric*.

We say that a set A is *chain-shrinkable* onto a subset B whenever every one of its chains is deformable onto B over A. We also say that B is an HR of A. The HR property is the analogue of the retract property, however, with the difference that in retraction of A onto B under a c.s.v.t. the set B remains fixed point for point, while no such condition is imposed under chain-shrinking. The difference is more apparent than real as shown by:

THEOREM II. *If c_p is chain-deformable onto a closed set A on $\mathfrak{R}$, the chain-deformation ϑ of c_p onto A may be so chosen as to be merely over $\overline{\mathfrak{R} - A}$. More precisely it may be associated with an ϵ-elementary decomposition, ϵ assigned, $\{c_q^i\}$ of c_p, such that the elements c_q^{2i} alone meet $\mathfrak{R} - A$ and that the others are not deformed: $\vartheta c_q^{2i+1} = c_q^{2i+1}$ ($\vartheta = 1$ on A).*

The basic step is the derivation of the elementary ϵ-decomposition. When the chains are singular we merely take a subdivision of mesh ϵ and call c_q^{2i} its elements on A and c_q^{2i+1} the rest. The problem is more difficult when the chains are regular.

Let first L be a fundamental complex of A which is also a subcomplex of K (L1, p. 474) and let there be given a chain-deformation ϑ of c_p associated with an elementary ϵ-decomposition $\mathfrak{K}' = \{c_q'^i\}$ and the deformation-chains $\mathfrak{D}' c_q^i$. If $C_{q+1}'^i$ is the representative of $c_q'^i$ in K and $c_q^i \not\subset A$, we may write

$$C_{q+1}'^{i} = C_{q+1}^{2i} + C_{q+1}^{2i+1}, \tag{9.1}$$

where $C_{q+1}^{2i} = L \cdot C_{q+1}'^{i}$ and C_{q+1}^{2i+1} is the remaining part. The two chains at the right of (9.1) determine chains c_q^{2i}, c_q^{2i+1} such that

$$c_q'^{i} = c_q^{2i} + c_q^{2i+1}, \tag{9.2}$$

where c_q^{2i+1} alone meets $\mathfrak{R} - A$. If $c_q^i \subset A$, we set $c_q^{2i} = c_q^i$, $c_q^{2i+1} = 0$. In any case c_q^{2i}, c_q^{2i+1} are both on $|c_q'^i|$, hence their diameters $< \epsilon$, so that $\mathfrak{K} = \{c_q^i\}$ is likewise an ϵ elementary decomposition of c_p. Moreover except for the existence of a suitable ϑ it behaves as demanded by the theorem. We shall now construct ϑ.

Regarding the operation $\mathfrak{D}$ to be associated with ϑ we first specify that $\mathfrak{D} c_q^{2i} = 0$. Next we treat $\mathfrak{D}' c_q^{2i+1}$ like $c_q'^i$ above and obtain

$$\mathfrak{D}' c_q^{2i+1} = d_{q+1}^i + d_{q+1}'^i, \tag{9.3}$$

where $d_{q+1}^i \subset A$, $d_{q+1}'^i \subset \overline{\mathfrak{R} - A}$, and we now set $\mathfrak{D} c_q^{2i+1} = d_{q+1}'^i$. We now determine the chain-decomposition $\{c_q''^i\}$ of the new transform chain ϑc_q by the boundary relations

$$\mathfrak{D} c_q \to \vartheta c_q^i - c_q^i - \mathfrak{D} F(c_q^i), \tag{9.4}$$

in which for every combination i, q, the only unknown term is $c_q''^i = \vartheta c_q^i$, which is thus uniquely determined by (9.4).

Since the chains $F(c^{2i})$ are of c^{2i} type, their $\mathfrak{D}$ in (9.4) are both zero, hence $c_q''^{2i} = c_q^{2i}$. Similarly we have $c_q''^{2i+1} \subset \overline{\mathfrak{R} - A}$. Finally,

$$\mathfrak{D}' c_q^i \to \vartheta' c_q^i - c_q^i - \mathfrak{D}' F c_q^i \tag{9.5}$$

for every value of i. Hence

$$(\mathfrak{D}' - \mathfrak{D})\, c_q^i \to (\vartheta' - \vartheta)\, c_q^i - (\mathfrak{D}' - \mathfrak{D})\, F c_q^i. \tag{9.6}$$

Since the chains $(\mathfrak{D}' - \mathfrak{D})\, c_q^i \subset A$, $\vartheta' c_q^i \subset A$, (9.6) shows that $\vartheta c_q^i \subset A$ also and hence $\vartheta\, c_p \subset A$. Therefore all the conditions of the theorem are effectively fulfilled.

REMARK. Since $|\, \mathfrak{D}\, c_q^i \,| \subset |\, \mathfrak{D}'\, c_q^i \,|$, if the initial chain-deformation ϑ' is ϵ, so is the modified one ϑ.

COROLLARY. *If A is open instead of closed, we may choose $\{c_q^i\}$ such that $c_q^{2i} \subset A, c_q^{2i+1} \subset \mathfrak{R} - A$, with $\vartheta\, c_q^{2i} = c_q^{2i}$ ($\vartheta = 1$ on A).*

For all that is necessary is to apply the theorem to A.

APPLICATION. *If $F(c_p) \subset A$ the chain-deformation may be so chosen as to leave $F(c_p)$ unchanged.*

10. The notion of chain-shrinking as here presented suffers from the disadvantage of not being *local*. The "local" properties are, however, frequently the most important, and so we shall consider them now.

Let A, B be subsets of $\mathfrak{R}$. We say that A may be *chain-shrunk away from B* whenever there is an open set $U \supset B$ such that A may be chain-shrunk onto $A - U$. We have at once

THEOREM III. *If A may be chain-shrunk away from every point of a compact set B, it may be shrunk away from B.*

By the Borel covering theorem, B has a f.c.o.s. $\{U^i\}$, such that A is shrinkable onto every $A - U^i$. We can find another f.c.o.s. $\{V^i\}$ of B such that $\bar{V}^i \subset U^i$.[11] We shall use both coverings simultaneously.

By assumption A may be chain-shrunk onto $B - U^1$ and as a consequence every chain c_p of A will have been displaced to the exterior of a certain open set $W^1 \supset \bar{V}^1 \cdot B = B^1$.

Suppose that we have succeeded in showing that the displacement may be carried out similarly to outside a certain open set

$$W^{k-1} \supset B^{k-1} = (\bar{V}^1 + \cdots + \bar{V}^{k-1}) \cdot B\,.$$

If $W^{k-1} \supset B$ we are through. In the contrary case, there is at least one more V, which we may assume to be V^k such that $\bar{V}^k \cdot B \not\subset W^{k-1}$. In any case however the chains of A not on W^{k-1} will be at a positive distance δ from B^{k-1}. By assumption we can displace all the chains of A onto $A - U^k$. However by Theorem II, this displacement may be replaced by one onto the exterior of the spherical neighborhood $\mathfrak{S}(U^k \cdot B, 1/2\, \delta)$. The chains thus displaced will be at a positive distance from B^k, hence outside of some $W^k \supset B^k$. Proceeding thus we shall ultimately have a $W^n \supset B$, with all chains deformable onto $A - W^n$. The theorem is therefore proved.

§4. HLC spaces and their relations to other spaces

11. The definition of the HLC sets of various types is entirely similar to that of LC sets. Here also, however, we need the *local* characterization.

[11] Menger, *Dimensionstheorie*, p. 160.

We shall say then that $\mathfrak{R}$ is:

p-HLC at the point x whenever every open set $U \supset x$ contains another $V \supset x$ also, such that every p-cycle, $p > 0$, of V is ~ 0 on U, and every zero-cycle γ_0 of V is $\sim$ on U to a point of V taken (γ_0) times;

HLC^p at x whenever it is q-HLC for every $q \leqq p$;

HLC$^\infty$ at x whenever it is q-HLC for every q whatsoever;

weak HLC at x whenever V can be determined as above independently of q.

The strong HLC (merely HLC) will be defined for a compact metric $\mathfrak{R}$ in No. 15.

Finally $\mathfrak{R}$ itself is p-HLC, $\cdots$ whenever it has the corresponding property at all its points.

The ϵ, η formulations in the compact metric case are as usual: in place of "for every U there is a V" we must have "for every $\epsilon > 0$ there is an $\eta > 0$". In particular the HLC^p condition may be reduced to "for every ϵ there is an $\eta(\epsilon, p)$ such that every q-cycle $q < p$ (a γ_0 whose $(\gamma_0) = 0$) of diameter $< \eta$ bounds a chain of diameter $< \epsilon$", while the weak HLC condition will be of the same form with η independent of p.

12. If c_p and c_{p-r}^j are two chains of a finite quasi-complex $\mathfrak{K}$, we shall say that they are *incident* whenever c_{p-r}^j is a boundary chain of c_p^i. The aggregate of the specifications of the incidences of $\mathfrak{K}$ is called the *pattern* of $\mathfrak{K}$.

Suppose that we have given a $\mathfrak{K}$ whose pattern we are to reproduce on $\mathfrak{R}$, and of all the elements expected let there be present all the zero-chains and some, but not necessarily all, of the rest, so that the chains present form a quasi-sub-complex $\mathfrak{K}'$ of $\mathfrak{K}$. We call $\mathfrak{K}'$ a *partial realization* of $\mathfrak{K}$. Let us suppose that out of an expected c_p^i and its boundary chains, there are already present the chains $c_r'^{\,i}$ in $\mathfrak{K}'$. We call max diam $\Sigma \mid c_r'^{\,i} \mid$ the *mesh* of the partial realization $\mathfrak{K}'$ of $\mathfrak{K}$.

If c_q^i is an expected chain of $\mathfrak{K}$ and $F(c_q^i)$ is already present in $\mathfrak{K}'$ a necessary condition is that it be a cycle. Furthermore if $q = 1$ and if its imposed boundary relations are

$$c_1^i \rightarrow \eta_{ij} c_0^j \,, \tag{12.1}$$

where, by assumption, the chains c_0^j are already present, from (12.1) follows

$$\eta_{ij} (c_0^j) = 0 \,, \tag{12.2}$$

and these relations must be satisfied if the data are consistent. They shall naturally be assumed if $\mathfrak{K}'$ is given, or must be verified whenever a $\mathfrak{K}'$ is constructed.

13. Theorem IV. *N.a.s.c. for a compact metric space $\mathfrak{R}$ to be* HLC^p *are that for every $\epsilon > 0$ there exist an $\eta(\epsilon, p) > 0$ such that every partial realization $\mathfrak{K}'$ on $\mathfrak{R}$ of a $\mathfrak{K}_{p+1}$ whose mesh $< \eta(\epsilon, p)$ may be completed to make up the expected $\mathfrak{K}_{p+1}$ of mesh ϵ.*[12]

[12] Compare with the analogous proposition for LC^p-sets, L2, p. 120.

Since $F(c_q)$, $q \leqq p$, is a special $\mathfrak{K}'$ corresponding to a $\mathfrak{K}$ which consists of c_q and $F(c_q)$, the condition of the theorem is clearly sufficient for an HLC^p. We must therefore merely show its necessity, and as it is trivial for $p = -1$, we may take $p > 0$ and use induction on p.

Under our assumptions the theorem holds for $p - 1$, and there is a corresponding $\eta(\epsilon, p - 1)$. Moreover since $\mathfrak{R}$ is p-HLC we have also the constant $\xi(\epsilon, p)$ of the p-HLC property ($\eta(\epsilon, p)$ in the definition).

Let the mesh α of $\mathfrak{K}'$ be $< \eta(\beta, p - 1) < \beta, \beta > 0$ assigned. Under the hypothesis of the induction we may insert all the missing chains of dimension $\leqq p$ and have their diameters $< \beta$. We thus obtain a new partial realization $\mathfrak{K}''$ of $\mathfrak{K}_{p+1}$ in which only the $(p + 1)$-chains may be missing. But if c^h_{p+1} is any expected chain, the sum of the sets of its boundary chains is of diameter $< \alpha + 2\beta < 3\beta$. Therefore, if we choose $\beta < \frac{1}{3}\, \xi(\epsilon, p)$, we may insert the missing chains and choose them of diameter $< \epsilon$. Hence the condition of the theorem is fulfilled whenever the mesh $\alpha < \eta\, (\frac{1}{3}\, \xi(\epsilon, p)), p - 1) = \eta(\epsilon, p)$.

14. Theorem V. *Given a compact metric* HLC^p *space* $\mathfrak{R}$ *there exists for every* $\epsilon > 0$ *a quasi-complex* Ψ_p *into a subchain of which every* c_q, $q \leqq p$, *of* $\mathfrak{R}$, *is* ϵ*-deformable over* $\mathfrak{R}$.

Let $\{U^i\}$ be a f.c.o.s. of the HLC^p space $\mathfrak{R}$ of Theorem IV, and let α be the mesh and Φ the skeleton of the covering. We shall assume that $\{\bar{U}^i\}$ has the same skeleton Φ, a choice always possible whatever α (see L1, p. 474). This means that any group of U's intersect when and only when their closures do. As to the latter, we know that there exists a constant $\gamma < \alpha$ such that whenever any set of diameter $< \gamma$ meets a group of U's, these U's intersect. From the preceding we conclude then that γ has also the same property as regards the U's themselves.

Let us mark on U^i a point A^i which we consider as an oriented zero-cell. If Φ_p is the sub-complex obtained after removing all simplexes of dimension $> p$ from Φ, we may consider the set $\{A^i\}$ as a partial realization Φ'_p of a quasi-complex Ψ_p with the same incidence pattern as Φ_p. It may likewise be considered as a partial realization Ψ' of a quasi-complex Ψ with the same pattern as Φ. From the construction of Φ we conclude that the meshes of the two partial realizations do not exceed the maximum diameter 2α of the sum of any group of intersecting U's. It follows in particular that if $\alpha < \frac{1}{2}\eta(\beta, p)$ we may complete Ψ'_p to a Ψ_p of mesh $< \beta$.

It is now a simple matter to show that if $\delta =$ mesh $\mathfrak{K}_p < \gamma$ then $\mathfrak{K}_p$ may be ϵ-chain deformed into a quasi-complex whose elementary chains are subchains of Ψ_p. In fact the chain-deformation ϑ may be so chosen that: (a) every elementary c^i_0 of $\mathfrak{K}_p$ goes into (c^i_0) times a vertex of $\mathfrak{K}_p$; (b) if c^j_q is an elementary chain of $\mathfrak{K}$ with zero-chains in its boundary then $\vartheta\, c^j_q$ is the image in the transformation $\Phi_p \to \Psi_p$ of a chain of Φ_p which is on the $F(\sigma)$ of the simplex σ whose vertices have for images the points A^i corresponding to the boundary zero-chains of c^j_q; (c) if c^j_q is not of type (b) then $\vartheta\, c^j_q = 0$.

We satisfy (a) by choosing for $\vartheta\, c_0^i$ a chain $(c_0^i)A^h$ where A^h is such that U^h meets c_0^i. By the same token (b) holds for $q = 0$, and (c) does not occur for that value. From this moment on the proof proceeds essentially like that of the deformation theorem of *Topology*, p. 92. The only point that may raise a question is the following. Granting that we have case (b) and that $\vartheta\, F(c_q^j)$, $q < p$, has already been described we must describe $\vartheta\, c_q^j$. Owing to $\delta < \gamma$ the vertices A^i to which $\vartheta\, F(c_q^j)$ belongs are on a set of intersecting U's, and hence they are the images of those of a σ of Φ. It follows that $\vartheta\, F(c_q^j)$ is the image of a cycle Γ_{q-1} of $F(\sigma)$. Now $\sigma + F(\sigma)$ has a subchain $C_q \to \Gamma_{q-1}$, and the image of C_q (under the chain-transformation $\Phi_p \to \Psi_p$) is a chain $c_q'^{\,i} \to F(c_q^i)$. We set $\vartheta\, c_q^i = c_q'^{\,i}$ and extend ϑ to all q-chains by the linearity condition. The rest of the proof is as loc. cit.

15. Paraphrasing the treatment of L2, No. 5, we now define a set as *strong* HLC or merely HLC when it is HLC^p for every p, and when in addition the function $\eta(\epsilon, p)$ of Theorem IV has a lower bound $\eta(\epsilon) > 0$, independent of p.

The results of our paper regarding the comparison with retracts may also be extended here, provided that absolute retraction is defined relatively to imbedding not in an arbitrary set but in an arbitrary HLC set. We shall not stop to develop this point further. From Theorem V follows readily:

THEOREM VI. *Every chain of an* HLC *is deformable into a subchain of a definite chain-complex.*

For in the present instance we may complete Ψ_p up to the dimension n of Φ, and have a complete chain-image Ψ of Φ. We then apply the proof of Theorem V with the supplementary condition that $\vartheta c_q^i = 0$ for $q > n$ which is consistent since $\Gamma_{q-1} = \Gamma_n$, being now a chain of a σ_n, will be $= 0$. It follows that every $\mathfrak{K}$ is deformable into a chain-complex whose elements of dimension $p \leqq n$ are *subchains* of Φ, and the others are zero.

THEOREM VII. *If* $\mathfrak{R}$ *is* HLC^n *and* $n = \dim \mathfrak{R}$ *is finite the space is* HLC.

For $\mathfrak{R}$ possesses no chains of dimension $> n$, hence in the construction connected with Theorem IV, one need never go beyond chains of dimension n and we may take for $\eta(\epsilon)$ the least of the $n + 1$ positive numbers

$$\eta(\epsilon, p),\ p = 0, 1, \cdots, n.$$

16. THEOREM VIII. *The homology groups of an* HLC^p, *absolute or mod a closed* HLC^p *subset, for dimensions* $\leqq p$, *or of an* HLC, *absolute or mod a closed* HLC *subset, all have the same structure as for a finite complex.*

By this we mean that they have finite bases with a finite number of relations between the elements of the bases, and furthermore for an HLC set that they are zero for dimensions above a certain integer n. In particular when the groups are free groups the Betti-numbers are all finite.

The proof is very simple. Take first the absolute groups and regular chains. The chains of $\vartheta\mathfrak{K}$ are the images in one of the correspondences $\Phi_p \to \Psi_p$, $\Phi \to \Psi$

of certain chains of Φ_p or Φ and the correspondence preserves boundary relations. Hence the groups $\{\gamma_q\}$ of $\mathfrak{R}$ ($q \leqq p$ for an HLC^p) are isomorphic with certain subgroups of the same for Φ_p or Φ respectively. They are therefore additive groups generated by a finite number of linear forms with coefficients in the abelian group $\mathfrak{M}$, corresponding say to chains γ_q^i, $i = 1, 2, \cdots, r$. The homologies between the γ's are those on $\mathfrak{R}$, all of the form

$$x_i \gamma_q^i \sim 0, \qquad x_i \subset \mathfrak{M}. \tag{16.1}$$

Since $\mathfrak{M}$ is a field this system may be reduced by a change of base to the form

$$\gamma_q^{s+i} \sim 0, \qquad i = 1, 2, \cdots, r - s \tag{16.2}$$

and the first s elements of the new base form a minimum base for the q^{th} homology group.

If we deal with singular chains, we may reason substantially as in the proof of the invariance of the combinatorial characters of a complex (*Topology*, p. 88) and show that (16.1) implies that there exists a chain of Ψ_p, or Ψ as the case may be,

$$c_{q+1} \rightarrow x_i \gamma_q^i . \tag{16.3}$$

Since the γ's may after all include all the q-subcycles of Ψ_p or Ψ, this shows that the q^{th} homology group is the same as for Ψ_p or for Ψ and hence the same as for Φ_p or for Φ, as the case may be.

17. Consider now a closed HLC^p subset G of an HLC^p space $\mathfrak{R}$ and let $\{U^i\}$ be the f.c.o.s. of No. 14. It may happen that certain sets U which meet G have an intersection which does not meet G. In that case their closure H also fails to meet G. Since the number of sets H is finite and they are closed, the distance δ of their sum from G is > 0. Let now, for every i, V^{2i} be the set of the points of U^i nearer than $2/3\ \delta$ from G, and V^{2i+1} the set of those farther than $1/3\ \delta$ from G. One of the two sets may well be vacuous. In any case however the non-vacuous sets V make up a new f.c.o.s. $\{V^i\}$ of $\mathfrak{R}$, which like the initial covering has the same incidence pattern as the covering $\bar{V}^i$ of the closures. Moreover, now if any aggregate of V's intersect G, they intersect on G itself. Let us assume then that the initial f.c.o.s. $\{U^i\}$ already possesses this property.

We now modify the construction of Ψ_p in No. 14, by choosing the point A^i on G whenever U^i meets G. Then, when the meshes are suitably small, we utilize the HLC^p property of G to place on that set every chain of Ψ_p whose vertices A^i are on G. This is always possible since, after all, the f.c.o.s. $\{U^i \cdot G\}$ behaves relatively to G like $\{U^i\}$ relatively to $\mathfrak{R}$ itself. As a consequence, $\vartheta\,\mathfrak{K}$ will transform a chain c_q^i of $\mathfrak{K}$ on G into chains of Ψ_p on G. Then, for suitably small meshes throughout, we shall be able to assign deformation-chains $\mathfrak{D}c_q^i$ likewise on G. Under the circumstances the chains on G will be merely deformed over G and hence the cycles mod G will be deformed into cycles of Ψ_p mod G. From this point on the rest of the proof is as for absolute cycles.

If $\mathfrak{R}$ and G are both HLC the treatment is the same except that Ψ is to take the place of Ψ_p throughout and the conclusion is again the same.

COROLLARIES. I. *If $\mathfrak{R}$ is* HLC^p *and G is* HLC^q, *$q < p$, the theorem regarding the homology groups of $\mathfrak{R}$ mod G holds for dimensions $\leqq q$.*

II. *If $n = \dim \mathfrak{R}$ is finite and both $\mathfrak{R}$ and G are* HLC^n, *the Betti-numbers absolute, or of $\mathfrak{R}$ mod G, if any occur, are all finite.*

APPLICATION. *If A is an* HLC^p*-set, $p \leqq n$, in an n-sphere H_n the Betti-numbers of $H_n - A$ are all finite; in particular R_0 is finite, and hence $H_n - A$ consists of a finite number of regions.*

For dim $A \leqq n$ and the rest follows from the preceding results together with the extension of Alexander's duality relation (*Topology*, p. 339).

18. **HNR-sets.** The closed set A shall be called an HNR of $\mathfrak{R}$ whenever for every positive ϵ there is a positive η such that every chain c_p within a spherical neighborhood $\mathfrak{S}(A, \eta)$ of A is ϵ-deformable onto A. These are obvious, but not complete, analogies of the NR property. The most important difference is that we have to use two neighborhoods of A, whereas in the NR there is only one.

THEOREM IX. *Let $\mathfrak{R}$ be compact, metric* HLC. *Then any closed subset A of $\mathfrak{R}$ which is also* HLC *is an* HNR *of the space.*

Let $\eta_1(\epsilon)$, $\eta_2(\epsilon)$ be the constants of the HLC definition relative to ϵ and respectively to $\mathfrak{R}$ and A. Let $c_p \subset \mathfrak{S}(A, \eta)$ and let us impose upon it an elementary decomposition of mesh ξ making up a quasi-complex $\mathfrak{K}$. For each c_0 of the subdivision we take a point B on A among those nearest to c_0. These points constitute a partial realization of mesh $< 2\eta + \xi$ of a $\mathfrak{K}'$ having the same structure as $\mathfrak{K}$. By the HLC property of A, if we have a given $\alpha > 0$ and if $2\eta + \xi < \eta_2(\alpha)$ we may complete the partial realization to form $\mathfrak{K}'$ of mesh $< \alpha$. Then $\mathfrak{K}' + \mathfrak{K}$ form a partial realization of a certain $\mathfrak{D}\mathfrak{K}$ of mesh $< \eta + \xi + \alpha$, and if this quantity $< \xi_1(\epsilon)$ we may complete to form $\mathfrak{D}\mathfrak{K}$ of mesh ϵ. As a consequence $\mathfrak{K}$, and hence c_p, will then have been ϵ-deformed onto A. The two required conditions may be fulfilled by choosing

$$\alpha < \tfrac{1}{3}\,\eta_1(\epsilon)\,; \qquad \xi, \eta < \tfrac{1}{3}\,\eta_2\left(\tfrac{1}{3}\,\eta_1(\epsilon)\right).$$

Therefore the theorem holds, with the $\eta(\epsilon)$ of the HNR definition here chosen as $\frac{1}{3}\,\eta_2\left(\frac{1}{3}\,\eta_1(\epsilon)\right)$.

REMARK. Here also as in Theorem II, the construction may be so modified that the elements of c_p, or rather of its elementary decomposition, already $\subset A$ are unmodified. For example, if $F(c_p) \subset A$ then it remains unmodified and $c_p' \sim c_p$.

19. **Relations between the classes LC^p, LC and HLC^p, HLC.**

THEOREM X. *In the system of singular chains an* LC^p *is an* HLC^p *and an* LC *is an* HLC.

The proof is by means of Theorem I of L2 (p. 120) which is the analogue for LC-sets of our present Theorem IV. Assuming then $\mathfrak{R}$ to be LC^p let $\xi(\epsilon, p)$

be the constant called $\eta(\epsilon, p)$ in L2, p. 121 (proof of Theorem I), the change of notation being in order to avoid confusing that constant with the $\eta(\epsilon, p)$ of Theorem IV. Let $\mathfrak{K}'_p$ be a partial realization of mesh $< \xi(\frac{1}{3}\epsilon, p)$ of a certain $\mathfrak{K}_p$. We may consider $\mathfrak{K}_p$ as the potential singular image of a certain simplicial complex K_p and the chains of $\mathfrak{K}'_p$ correspond to certain chains of K whose simplexes make up a subcomplex K' of K_p. We may also consider $\mathfrak{K}'_p$ as a singular image of certain chains and cycles of K'. By subdividing the chains of $\mathfrak{K}'_p$ and inserting suitable vertices on the chains already present we turn it into a singular image of the oriented simplexes which make up the chains in question in K. Now let c_1^i be any expected chain of $\mathfrak{K}_p$ and let C_0^i be the chain of K' whose image, already present, is $F(c_1^i)$. We build up a new K_p out of K' by taking the join of some point P^i with C_0^i and adding this new simplicial chain to K'. We then proceed similarly by adding new two-chains for any c_2^i whose boundary is now already represented in the complex already obtained, and so on until we have a new K_p which may manifestly replace the former.

Now let c_q^i be an expected chain of $\mathfrak{K}_p$ whose boundary is represented by certain chains in $\mathfrak{K}'$ and which has necessitated the addition of a certain number of vertices P^j in building up the new K_p. We represent these vertices by certain points on $F(c_q^i)$, and if there are chains wholly unrepresented in the process we represent them by zero. We have now in the modified $\mathfrak{K}'$ a semi-singular image of K_p in the sense of L2, and hence we may complete it to form a singular image on $\mathfrak{R}$ whose mesh $< \epsilon/3$, and which contains a collection of chains making up a chain-image of $\mathfrak{K}_p$. The chains are now singular with cells of diameters $< \xi(\epsilon/3, p) < \epsilon/3$, or else directly of diameters $< \epsilon/3$, and if c_q^i is new, its singular cells meet the boundary chains of c_q^i already present which are on a set of diameter $< \xi(\epsilon/3, p)$. Therefore mesh $\mathfrak{K} < \epsilon$ and by Theorem IV, $\mathfrak{R}$ is HLC^p, with $\eta(\epsilon, p) = \xi(\epsilon/3, p)$.

If $\mathfrak{R}$ is LC the constant ξ has a positive lower bound $\xi(\epsilon)$ independent of p, hence $\eta(\epsilon, p) \geqq \xi(\epsilon)$ and $\mathfrak{R}$ is HLC. Our theorem is therefore proved.

20. **Locally contractible spaces.** According to Borsuk to whom this notion is due,[13] a topological space $\mathfrak{R}$ is called locally contractible whenever for every

[13] See K. Borsuk, *Über eine Klasse von lokal zusammenhängenden Räumen*, Fundamenta Mathematicae, vol. 19 (1932), pp. 220–242, notably p. 236.

Borsuk has shown (loc. cit., p. 240) that if dim $\mathfrak{R} = n$ is finite, we have in an obvious sense ANR $\sim$ LC$^\infty$. He then states as open (footnote 39) the question of the equivalence of the two properties when dim $\mathfrak{R}$ is infinite. Now we have shown (L2, p. 121) that ANR $\sim$ LC and, as we shall show by an example, LC $\nsim$ LC$^\infty$, hence Borsuk's question must be answered negatively. Thus our criterion for ANR is definitely the only one which holds independently of the dimension. This emphasizes once more the importance of the *semi-singular* complex introduced in L2, and of its analogue, the *partial realization* of chain-complexes of the present paper.

The example alluded to above was communicated to us by Borsuk and is as follows. Consider in the Hilbert parallelotope a sequence of points $\{x_p\}$ on a segment Ox_1 from the origin O, $\to O$ monotonely, and let H_p be a euclidean p-sphere of center x_p and radius

open set U containing any point x there is another, $V \supset U$, which is homotopic to a point on U. Let $\mathfrak{R}$ be compact, metric, locally contractible and consider the class of all spheroids $\{\mathfrak{S}(x, \epsilon)\}$ of $\mathfrak{R}$ with fixed radius ϵ. For every point x there is a spheroid $\mathfrak{S}(x, \xi)$ which is ϵ-homotopic to a point. As the space $\mathfrak{R}$ is compact it has a f.c.o.s. $\{\mathfrak{S}(x_i, \xi)\}$ and there is a constant η such that every subset of $\mathfrak{R}$ whose diameter $< \eta$ is on one of the spheroids $\mathfrak{S}(x_i, \xi)$ and hence ϵ-homotopic to a point on $\mathfrak{R}$.

It follows immediately from what precedes that given ϵ every singular p-sphere of $\mathfrak{R}$ whose diameter $< \eta$ bounds a singular $(p + 1)$-cell of diameter $< \epsilon$, where $\eta(\epsilon)$ is independent of p. We might call this property weak LC.[14] Owing to Theorem X this implies that $\mathfrak{R}$ is a weak HLC in the system of singular chains. That it is a weak HLC in any system is readily shown as follows. Any cycle γ_p such that diam $|\gamma_p| < \eta$ is ϵ-chain deformable, by virtue of Theorem I, into a point, which is a $\gamma_p \sim 0$, when $p > 0$, and which is taken (γ_0) times when $p = 0$. Therefore (No. 11) $\mathfrak{R}$ is p-HLC for every p, and since we have been able to choose $\eta(\epsilon)$ independently of p, we have

THEOREM XI. *A locally contractible compact metric space is both a weak* LC *and a weak* HLC.

While local contractibility is not sufficient to insure the LC or HLC properties, this will be the case if we strengthen it in the following manner. Let us call the open set U *self-contractible* whenever it is homotopic to a point on itself, i.e., whenever it is its own set V in the definition of local contractibility. A class of open sets $\{U\}$ will be called *self-contractible* whenever it consists of self-contractible sets and is closed with respect to intersection (if U', U'' are in the class so is $U' \cdot U''$). *A compact metric space $\mathfrak{R}$ which possesses for every ϵ a self-contractible class whose sets are all of diameters $< \epsilon$, is both strong* LC *and strong* HLC. The proof is essentially parallel to the analogue in L2, pp. 126, 127 (No. 17), and need not be repeated here.

21. **Application to locally polyhedral spaces.** We shall say that $\mathfrak{R}$ is locally polyhedral whenever every point x has a neighborhood whose closure is a finite complex K^x. We may replace K^x by a subdivision with x as a vertex and then the neighborhood by the star of x in K^x. Therefore $\mathfrak{R}$ may be defined as a space of which every point has for neighborhood the star of a vertex in a finite complex. Since every star is homotopic on itself to its own vertex, the space $\mathfrak{R}$ is locally contractible and it is also clearly locally compact.

Suppose now $\mathfrak{R}$ to be also compact and metric. Since it may then be covered

$< \frac{1}{2} d(x_p, x_{p-1} + x_{p+1})$. $\mathfrak{R}$ is the sum of the spheres $+ O +$ the segment with the open diameters of the spheres on the segment removed. It is readily shown that the p-LC condition holds for every ϵ with the corresponding $\eta(\epsilon, p) <$ diam H_p. Hence $\eta(\epsilon, p) \to 0$ with increasing p and $\mathfrak{R}$ is LC$^\infty$ but not LC.

[14] Is the converse true, i.e., is every weak LC locally contractible? Is weak LC $\sim$ LC? For the present these questions must remain unanswered.

with a finite number of the stars, its dimension is finite. Combining this with No. 17 Corollary II and Theorem XI, we have then

THEOREM XII. *A locally polyhedral compact metric space is both* LC *and* HLC.

COROLLARY. *The homology groups of the space of Theorem* XI, *absolute or mod a closed subset of same type, have the structure of those of a finite complex. In particular its Betti-numbers, if any occur, are also finite.*

IMPORTANT SPECIAL CASE: The space and its subset are *absolute topological manifolds.*

Of course in both theorem and corollary the space is assumed compact.

22. **Extension to locally compact separable spaces.** Substantially all the results of the present section may also be extended to these spaces. We recall that any locally compact separable space is also metric, although the reverse need not hold.[15] As far as our definitions are concerned, the HLC^p or HLC conditions formulated in No. 11 in terms of ϵ, η must now refer to the cycles meeting a specific self-compact subset A of $\mathfrak{R}$, and must hold for every such A. However $\eta(A, \epsilon, p)$ or $\eta(A, \epsilon)$ may well vary with A. The U, V conditions applied to every point of A may be replaced in fact by ϵ, η conditions on the chains meeting A, but not on all chains. Keeping this in mind we verify at once that Theorem IV holds when the elementary chains of $\mathfrak{K}'$ all meet A. In Theorem V the chains considered must now all be chains of A and the deformation is on any preassigned open set $U \supset A$. The subsequent results on the homology groups hold regarding the homologies between chains of A on U, that is to say, when bounding is allowed not merely on A but also on U. Theorem X may be derived, provided that we extend, as we may readily do, Theorem I of L2 to locally compact separable spaces. It is always understood that the semi-singular complex loc. cit. must have all its (realized) elements meeting A. Similarly Theorem XI holds also, it being always understood that the constants ϵ, η are related to A as described above. Since Theorem XII is a mere corollary of Theorem XI, it is likewise applicable to locally compact separable spaces.

23. **Concluding remarks.** The theory developed in the preceding pages was not intended to be exhaustive, but only to cover enough ground for some fairly immediate applications. Thus by means of quasi-complexes mod A, we could introduce chain-deformations mod A and extend in an obvious manner the results of the present paper. As already observed, another extension would be to non-compact spaces, say to the very general spaces whose homology theory has been developed in recent years by Čech. These extensions however would be strictly mechanical and may safely be left to any experienced reader.

PRINCETON, N. J.

[15] See Alexandroff-Urysohn, *Mémoire sur les espaces topologiques compacts*, Verhandelingen der Koninklijke Akademie van Wetenschappen te Amsterdam, afdeeling natuurkunde, deel XIV, No. 1 (1929), p. 83.

ON CONVERGENCE IN LENGTH

By C. R. Adams and Hans Lewy

1. **Introduction.** Adams and Clarkson[1] have recently considered a sequence of functions $f_n(x)$ $(n = 1, 2, 3, \cdots)$ defined on an interval[2] (a, b) and subject to the following conditions: (i) $f_n(x)$ tends to a limit function $f_0(x)$ of bounded variation; (ii) the total variation $T_a^b(f_n)$ of $f_n(x)$ on (a, b) tends to the total variation $T_a^b(f_0)$ of $f_0(x)$ on (a, b). The notation $f_n(x) - v \rightarrow f_0(x)$ was employed to describe this situation, the symbol $- v \rightarrow$ being read "converges in variation." Under these conditions each curve $y = f_n(x)$ for n sufficiently large has a finite length in the sense of Peano;[3] and the pair of conditions, (i) and (ii′) $L_a^b(f_n) \rightarrow L_a^b(f_0)$, define what may be called "convergence in length", the notation $f_n(x) - l \rightarrow f_0(x)$ being used for brevity.[4]

The reader of AC may very naturally raise a question as to the relation between convergence in variation and convergence in length; to compare these notions is one of the objects of the present note.

Definition of notation. It will be convenient usually to designate by S, with or without a subscript, a set of points $a = x_0, x_1, \cdots, x_p = b$, with

$$x_0 < x_1 < \cdots < x_p.$$

In general B will stand for a broken line inscribed in a curve $y = f(x)$ and consisting of p segments, the i^{th} segment $(i = 1, 2, \cdots, p)$ having end-points at $(x_{i-1}, f(x_{i-1}))$, $(x_i, f(x_i))$ and length b_i. The function whose graph is B will frequently be referred to as $B(x)$. Broken lines inscribed in two distinct curves and determined by the same set S will be called "corresponding".[5]

Received February 8, 1935; presented to the American Mathematical Society, February 23, 1935.

[1] C. R. Adams and J. A. Clarkson, *On convergence in variation*, Bulletin of the American Mathematical Society, vol. 40 (1934), pp. 413–417; this paper will be referred to as AC.

[2] The *closed interval* is always to be understood.

[3] According to Peano, a curve (continuous or not) is said to have finite length if the lengths of all inscribed broken lines B are bounded; by definition their least upper bound is the length of the curve.

[4] S. Bochner has imposed the condition of convergence in variation ("abgeschlossene Konvergenz") to permit passage to the limit in a Stieltjes integral: *Monotone Funktionen, Stieltjessche Integrale und harmonische Analyse*, Mathematische Annalen, vol. 108 (1933), pp. 378–410. This condition and that of convergence in length are sometimes employed in the calculus of variations; see Tonelli, *Fondamenti di Calcolo delle Variazioni*, vol. 1, Bologna, 1922.

[5] Incidentally we shall nowhere employ the fact that an absolutely continuous function is the integral of its own derivative. On the other hand, some of the results herein contained can be used as a basis for the deduction of theorems concerning summable functions; for instance, the possibility of approximating in the mean a summable function by step-functions may be concluded from Theorem 4, and the fact that an absolutely continuous function is the integral of a summable function follows easily.

2. Preliminary remarks. Since L and T for a function f are defined respectively as the least upper bound of L and of T for broken lines inscribed in f, we have at once the fundamental *semi-continuity relations:* that $f_n(x) \to f_0(x)$ on (a, b) implies

$$(1) \qquad \underline{\lim}_{n\to\infty} L_a^b(f_n) \geqq L_a^b(f_0), \qquad \underline{\lim}_{n\to\infty} T_a^b(f_n) \geqq T_a^b(f_0).$$

Hence the relation $f_n(x) - l \to f_0(x)$ $[f_n(x) - v \to f_0(x)]$ on (a, b) implies that relation for every subinterval. Moreover it may be observed that the relations $f_n(x) \to f_0(x)$ on a set of points everywhere dense in (a, b) and $L_a^b(f_n) \to L_a^b(f_0)$ $[T_a^b(f_n) \to T_a^b(f_0)]$, *when* $f_0(x)$ *is continuous on* (a, b), imply $f_n(x) - l \to f_0(x)$ $[f_n(x) - v \to f_0(x)]$ on (a, b).

If $f(x)$ is of bounded variation on (a, b), we have

$$(2) \qquad b - a + T_a^b(f) \geqq L_a^b(f) \geqq \{(b - a)^2 + [T_a^b(f)]^2\}^{1/2},$$

both of the inequalities flowing at once from the definitions of L and T. From (2) it follows (i) that *each of the relations* $f_n(x) - l \to c$, $f_n(x) - v \to c$ (c = *constant*) *on* (a, b) *implies the other;* and (ii) that $L_a^x(f)$ is continuous on the left [right] at each point x where f has this property. If $f(x)$ is absolutely continuous on (a, b), both $T_a^x(f)$ and $L_a^x(f)$ are likewise.

Whether f (of bounded variation on (a, b)) is continuous or not, there always exists a sequence of inscribed broken lines B_n $(n = 1, 2, 3, \cdots)$ such that we have both $B_n - l \to f$ and $B_n - v \to f$. For the set S_n determining B_n can be so chosen that S_{n+1} includes S_n for each n, and $S = \sum_{n=1}^{\infty} S_n$ is everywhere dense in (a, b) and contains all the points of discontinuity of f. Then $B_n(x)$ converges to $f(x)$ on (a, b): at each point ξ' in S the convergence is evident, while at a point ξ not in S we have

$$\overline{\lim_{n\to\infty}} \mid f(\xi) - B_n(\xi) \mid \leqq T_{\xi'}^{\xi}(f) + \overline{\lim_{n\to\infty}} \mid f(\xi') - B_n(\xi') \mid + \overline{\lim_{n\to\infty}} T_{\xi'}^{\xi}(B_n)$$
$$\leqq 2\, T_{\xi'}^{\xi''}(f)$$

for $\xi' < \xi < \xi''$ and ξ', ξ'' in S; since f is continuous at ξ, the right-hand member tends to zero with $\xi'' - \xi'$. From (1) and the definitions of L and T follow the relations $L_a^b(B_n) \to L_a^b(f)$ and $T_a^b(B_n) \to T_a^b(f)$.

3. We first prove

THEOREM 1. *The relation* $f_n(x) - l \to f_0(x)$ *on* (a, b) *implies* $f_n(x) - v \to f_0(x)$; *the converse is not true even when* $f_n(x)$ $(n = 0, 1, 2, \cdots)$ *is assumed absolutely continuous.*

Any ϵ (> 0) being given, let B be a broken line inscribed in f_0 and satisfying the inequality $L_a^b(B) > L_a^b(f_0) - \epsilon$. For convenience we set

$$L_{x_{i-1}}^{x_i}(f_0) - b_i = \epsilon_i, \qquad x_i - x_{i-1} = d_i;$$

then $\Sigma\, \epsilon_i$ is clearly $< \epsilon$ and there exists an m such that we have

$$L_{x_{i-1}}^{x_i}(f_n) \leqq L_{x_{i-1}}^{x_i}(f_0) + \epsilon_i \qquad (i = 1, 2, \cdots, p;\, n > m).$$

From (2) we thus obtain for each i

$$\begin{aligned} T_{x_{i-1}}^{x_i}(f_n) &\leqq \left\{\left[L_{x_{i-1}}^{x_i}(f_n)\right]^2 - d_i^2\right\}^{1/2} \\ &\leqq [(b_i + 2\epsilon_i)^2 - d_i^2]^{1/2} \\ &\leqq (b_i^2 - d_i^2)^{1/2} + 2(b_i\epsilon_i)^{1/2} + 2\epsilon_i, \end{aligned}$$

and hence, by aid of Schwarz's inequality,

$$\begin{aligned} T_a^b(f_n) &\leqq T_a^b(f_0) + 2\left(\sum_{i=1}^{p} b_i \sum_{i=1}^{p} \epsilon_i\right)^{1/2} + 2 \sum_{i=1}^{p} \epsilon_i \\ &\leqq T_a^b(f_0) + 2\,[\epsilon L_a^b(f_0)]^{1/2} + 2\epsilon. \end{aligned}$$

In view of (1) the direct statement in our theorem is now proved. An example illustrating the failure of the converse, even when all the functions involved are absolutely continuous, is provided by the sequence $f_n(x)$ defined in §5 of AC.

Corollary.[6] $L_a^b\,[f(x)] = L_a^b\,[T_a^x(f)]$.

From (2) it follows that neither or both of these quantities are finite. In the second case the relation

$$|f(x_i) - f(x_{i-1})| \leqq T_a^{x_i}(f) - T_a^{x_{i-1}}(f)$$

shows that, of two corresponding broken lines inscribed in $f(x)$ and $T_a^x(f)$, the former has length no greater than the latter. Hence $L_a^b[f(x)]$ is $\leqq L_a^b[T_a^x(f)]$. On the other hand, for any broken line B we have $L_a^b(B) = L_a^b[T_a^x(B)]$; hence, choosing a sequence of broken lines B_n $(n = 1, 2, 3, \cdots)$ with $B_n - l \rightarrow f$ we obtain

$$L_a^b(f) = \lim_{n\to\infty} L_a^b(B_n) = \lim_{n\to\infty} L_a^b\,[T_a^x(B_n)] \geqq L_a^b\,[T_a^x(f)],$$

since by Theorem 1 $T_a^x(B_n)$ tends to $T_a^x(f)$.

By virtue of Theorem 1 and the results obtained in §3 of AC, and by reasoning analogous to that employed in §3 of AC, we may readily establish the following theorems.

Theorem 2. *The relations $f_n(x) - l \rightarrow f_0(x)$ on (a, b) and $f_0(x') = f_0(x' - 0)$ $[f_0(x') = f_0(x' + 0)]$ imply that x' is a point of uniform convergence on the left [right] for $f_n(x)$, $T_a^x(f_n)$, and $L_a^x(f_n)$.*

Corollary. *The relation $f_n(x) - l \rightarrow f_0(x)$ on (a, b), with $f_0(x)$ continuous, implies that the convergence of $f_n(x)$ to $f_0(x)$, of $T_a^x(f_n)$ to $T_a^x(f_0)$, and of $L_a^x(f_n)$ to $L_a^x(f_0)$ is uniform over (a, b).*

Theorem 3. *The relations $f_n(x) - l \rightarrow f_0(x)$ on (a, b) and $f_0(x') \neq f_0(x' - 0)$*

[6] Dr. J. A. Clarkson suggested this relation and gave a proof.

$[f_0(x') \neq f_0(x' + 0)]$ *imply that* x' *is a point of uniform convergence on the left* [*right*] *for* $f_n(x)$, $T_a^x(f_n)$, *and* $L_a^x(f_n)$ *or for none of the three, according as* $f_n(x' - 0)$ $[f_n(x' + 0)]$ *tends to* $f_0(x' - 0)$ $[f_0(x' + 0)]$ *or not.*

COROLLARY. *If we have* $f_n(x) - l \to f_0(x)$ *on* (a, b), *a necessary and sufficient condition that* $T_a^x(f_n) \to T_a^x(f_0)$ *uniformly on* (a, b), *or that* $L_a^x(f_n) \to L_a^x(f_0)$ *uniformly on* (a, b), *is that* $f_n(x) \to f_0(x)$ *uniformly on* (a, b).

4. **Further inequalities.** If $f(x)$ and $g(x)$ are of bounded variation on (a, b), we have

$$T_a^b(f) + T_a^b(g) \geqq T_a^b(f + g); \tag{3}$$

$$L_a^b(f) + T_a^b(g) \geqq L_a^b(f + g); \tag{4}$$

$$M_2T_a^b(f) + M_1T_a^b(g) \geqq T_a^b(f \cdot g), \tag{5}$$

where $M_1 = \text{l.u.b.} \mid f \mid$ and $M_2 = \text{l.u.b.} \mid g \mid$;

$$[L_a^b(f)]^2 - (b - a)^2 (1 + m^2) \geqq [T_a^b(f - mx)]^2/(1 + m^2), \tag{6}$$

where $m = [f(b) - f(a)]/(b - a)$.

Relation (3) is well known and obvious; (5) is an immediate consequence of the definition of T. Relation (4) follows at once from the fact that the i^{th} segment of a broken line inscribed in $f + g$ has length no greater than the corresponding segment inscribed in f plus the quantity $\mid g(x_i) - g(x_{i-1}) \mid$.

Proof of (6). We first establish the inequality for the case of a broken line B, denoting the derivative $B'(x)$ by $\varphi(x)$ so that $B(x) = \int_a^x \varphi \, dx$. Then (6) reduces to[7]

$$\left[\int_a^b (1 + \varphi^2)^{1/2} \, dx\right]^2 \geqq (b - a)^2(1 + m^2) + \left[\int_a^b \mid \varphi - m \mid dx\right]^2 \Big/ (1 + m^2), \tag{7}$$

where $m = \int_a^b \varphi \, dx/(b - a)$. Let us designate by $\int_+ \left[\int_-\right]$ an integral over the set of points x for which $\varphi - m$ is $\geqq 0$ $[< 0]$. By Minkowski's inequality we have

$$\int_+ (1 + \varphi^2)^{1/2} \, dx \geqq \left[\left(\int_+ dx\right)^2 + \left(\int_+ \varphi \, dx\right)^2\right]^{1/2},$$

together with the same relation for $\int_-$. Setting

$$\alpha = \int_+ dx, \qquad \beta = \int_+ \varphi \, dx, \qquad \gamma = \int_- dx, \qquad \delta = \int_- \varphi \, dx,$$

[7] It is of interest to note that, as soon as (6) is proved, (7) holds for any summable function φ.

we obtain $m = (\beta + \delta)/(\alpha + \gamma)$ and

$$\int_a^b |\varphi - m|\,dx = \int_+ (\varphi - m)\,dx - \int_- (\varphi - m)dx$$
$$= \beta - m\alpha - \delta + m\gamma$$
$$= 2(\beta\gamma - \alpha\delta)/(\alpha + \gamma)\,,$$

and (7) will be true if

$$(8)\quad |\alpha + i\beta| + |\gamma + i\delta| \geqq \left[(\alpha + \gamma)^2 + (\beta + \delta)^2 + \frac{4(\beta\gamma - \alpha\delta)^2}{(\alpha + \gamma)^2 + (\beta + \delta)^2}\right]^{1/2}$$

holds. But under a suitable orthogonal transformation of a ξ, η-plane, the two points $\xi = \alpha, \eta = \beta$ and $\xi = \gamma, \eta = \delta$ respectively go into (α', β') and (γ', δ') with $\beta' + \delta' = 0$, while (8) is invariant and reduces to the triangular inequality

$$|\alpha' + i\beta'| + |\gamma' - i\beta'| \geqq |\alpha' + \gamma' + 2i\beta'|.$$

Relation (7) having now been proved, (6) follows at once by approximating f by a sequence of inscribed broken lines B_n $(n = 1, 2, 3, \cdots)$ with $B_n - l \rightarrow f$ and taking account of (1).

5. **Addition and multiplication of sequences.** In this section we examine the question of invariance under addition and multiplication of the property of convergence in length. That *convergence in variation is not invariant under these operations, even when all the functions involved are assumed absolutely continuous*, has been shown in AC. That *convergence in length is not invariant under addition, even when the limit functions are assumed continuous and all the approximating functions absolutely continuous*, may be seen from the following example.

Let $f_0(x)$ be defined on (0, 1) as the Cantor function, continuous but not absolutely continuous,[8] and let $f_0(x) = 1$ for $1 \leqq x \leqq 2$. We first observe that, if B is any broken line inscribed in f_0 on (0, 1), we have

$$T_0^1(f_0 - B) = \int_0^1 |B'|\,dx + T_0^1(f_0) = 2;$$

this shows that $f_n - l \rightarrow f_0$ on (0, 1), with f_0 continuous and f_n $(n = 1, 2, 3, \cdots)$ absolutely continuous, does not imply $T_0^1(f_n - f_0) \rightarrow 0$. In view of a theorem of Plessner[9] there exists a $\beta > 0$ and a sequence of positive numbers α_n $(n = 1, 2, 3, \cdots)$ such that $\alpha_n \rightarrow 0$ as $n \rightarrow \infty$ and $T_0^1[f_0(x) - f_0(x + \alpha_n)]$ is $> \beta$ for each n. Let a sequence of sets S_n $(n = 1, 2, 3, \cdots)$ on (0, 1) be chosen as

[8] Any other function of this character will do equally well.

[9] Plessner, *Eine Kennzeichnung der totalstetigen Funktionen*, Crelle's Journal, vol. 160 (1929), pp. 26–32; see also Dunford, *On a theorem of Plessner*, to appear in the Bulletin of the American Mathematical Society, vol. 41 (1935).

follows: for each n let the set S'_n determine a sum approximating T_0^1 $[f_0(x) - f_0(x + \alpha_n)]$ within $\beta/2$; let sequence S_n be selected so that S_n includes $S_{n-1} + S'_n$ for each n and $S = \sum_{n=1}^{\infty} S_n$ is everywhere dense in (0, 1). Designate by B_n, B^*_n, and C_n respectively the broken lines inscribed in $f_0(x)$, $f_0(x + \alpha_n)$, and $f_0(x) - f_0(x + \alpha_n)$ and determined by S_n. We then have $B_n(x) - B^*_n(x) \equiv C_n(x)$, $B^*_n(x) - f_0(x) \equiv B_n(x) - C_n(x) - f_0(x)$, $B_n(x) \to f_0(x)$, and $| C_n(x) | \leqq \max | f_0(x) - f_0(x + \alpha_n) | \to 0$, whence $B^*_n(x) \to f_0(x)$. Moreover we have by (1)

$$L_0^1 [f_0(x)] \leqq \varliminf_{n\to\infty} L_0^1(B^*_n) \leqq \varlimsup_{n\to\infty} L_0^1(B^*_n) \leqq \varlimsup_{n\to\infty} L_0^1 [f_0(x + \alpha_n)] = L_0^1 [f_0(x)].$$

Hence we have $B^*_n - l \to f_0$. We also have $B_n - l \to f_0$. On the other hand $T_0^1(B_n - B^*_n) = T_0^1(C_n)$ is $> \beta/2$ for each n, which implies (see Theorem 1) that $B_n(x) - B^*_n(x)$ does not converge in length to the function zero.

If, however, the limit functions are *absolutely* continuous, the property of convergence in length is invariant under addition and multiplication independently of the nature of the approximating functions. To establish this fact we first prove[10]

THEOREM 4. *The relation $f_n(x) - l \to f_0(x)$ on (a, b), when $f_0(x)$ is absolutely continuous, implies $T_a^b(f_n - f_0) \to 0$.*

That the hypothesis of absolute continuity for f_0 cannot be dispensed with is shown by the example just described above. For the proof of the theorem we employ two lemmas concerning an absolutely continuous function f_0, as follows.

a) If f_0 is absolutely continuous, S an arbitrary set subdividing (a, b), B_0 the broken line inscribed in f_0 and determined by S, and $\mathfrak{J}_h$ $(h = 1, 2, \cdots, q)$ those subintervals in which the slope m_h of B_0 is in absolute value $\geqq M$, there exists a function $\alpha(M)$ independent of S such that

$$\sum_h T_{\mathfrak{J}_h}(f_0) \leqq \alpha(M), \qquad \text{and} \qquad \alpha(M) \to 0 \qquad \text{as} \qquad M \to \infty.$$

To prove this assertion, we observe first $\Sigma\mathfrak{J}_h M \leqq \Sigma\mathfrak{J}_h | m_h | \leqq T_a^b(f_0)$, whence $\Sigma\mathfrak{J}_h \leqq T_a^b(f_0)/M$. Then, on account of the absolute continuity of f_0, $T_a^x(f_0)$ is absolutely continuous and for any set of non-overlapping intervals the sum of whose lengths is $\leqq T_a^b(f_0)/M$, and in particular for our set of intervals $\mathfrak{J}_h$, we have

$$\Sigma T_{\mathfrak{J}_h}(f_0) \leqq \alpha(M),$$

with $\alpha(M) \to 0$ as $M \to \infty$.

b) The hypotheses of a), together with $f_n(x) - v \to f_0(x)$ on (a, b), imply

$$\varlimsup_{n\to\infty} \sum_h T_{\mathfrak{J}_h}(f_n - f_0) \leqq \varlimsup_{n\to\infty} \sum_h T_{\mathfrak{J}_h}(f_n) + \sum_h T_{\mathfrak{J}_h}(f_0) \leqq 2\alpha(M).$$

[10] A proof has been given by Tonelli under the additional assumption that all the functions $f_n(x)$ $(n = 1, 2, 3, \cdots)$ are absolutely continuous; loc. cit., pp. 186–187.

The conclusion is immediate, since only a finite number of intervals $\mathfrak{J}_h$ is involved and in each we have $T(f_n) \to T(f_0)$.

Turning now to the proof of Theorem 4, let S be any set on (a, b) determining a broken line B_n inscribed in f_n $(n = 0, 1, 2, \cdots)$; for each subinterval defined by S we clearly have

$$\overline{\lim_{n\to\infty}} T_{x_{i-1}}^{x_i}(f_n - f_0) \leqq \overline{\lim_{n\to\infty}} T_{x_{i-1}}^{x_i}(f_n - B_n) + \overline{\lim_{n\to\infty}} T_{x_{i-1}}^{x_i}(B_n - B_0) + T_{x_{i-1}}^{x_i}(f_0 - B_0)$$

$$\leqq \overline{\lim_{n\to\infty}} T_{x_{i-1}}^{x_i}(f_n - B_n) + T_{x_{i-1}}^{x_i}(f_0 - B_0).$$

For those intervals $\mathfrak{J}_h$ for which the slope m_h of B_0 is numerically $\geqq M$, we have by Theorem 1 and lemma b)

$$\overline{\lim_{n\to\infty}} \Sigma\, T_{\mathfrak{J}_h}(f_n - f_0) \leqq 2\alpha(M), \qquad \alpha(M) \to 0 \text{ as } M \to \infty. \tag{9}$$

Denoting by Σ' the summation over only those intervals of S for which the slope m_h of B_0 is numerically $< M$, we obtain by (6) and Schwarz's inequality

$$\Sigma' T_{x_{i-1}}^{x_i}(f_0 - B_0) \leqq (1 + M^2)^{1/2}\, \Sigma' \left\{\left[L_{x_{i-1}}^{x_i}(f_0)\right]^2 - \left[L_{x_{i-1}}^{x_i}(B_0)\right]^2\right\}^{1/2}$$

$$\leqq (1 + M^2)^{1/2}\Sigma'\left[L_{x_{i-1}}^{x_i}(f_0) + L_{x_{i-1}}^{x_i}(B_0)\right]^{1/2}\left[L_{x_{i-1}}^{x_i}(f_0) - L_{x_{i-1}}^{x_i}(B_0)\right]^{1/2} \tag{10}$$

$$\leqq (1 + M^2)^{1/2} K(B_0),$$

where

$$K(B_0) = [L_a^b(f_0) + L_a^b(B_0)]^{1/2}\,[L_a^b(f_0) - L_a^b(B_0)]^{1/2}.$$

Similarly, on summing over these *same* intervals, we find in view of the convergence of B_n to B_0

$$\overline{\lim_{n\to\infty}} \Sigma' T_{x_{i-1}}^{x_i}(f_n - B_n) \leqq (1 + M^2)^{1/2}\, \overline{\lim_{n\to\infty}}\, [L_a^b(f_n) + L_a^b(B_n)]^{1/2}[L_a^b(f_n) - L_a^b(B_n)]^{1/2}.$$

Hence from $L_a^b(f_n) \to L_a^b(f_0)$ and $L_a^b(B_n) \to L_a^b(B_0)$ we infer

$$\overline{\lim_{n\to\infty}} \Sigma' T_{x_{i-1}}^{x_i}(f_n - B_n) \leqq (1 + M^2)^{1/2} K(B_0). \tag{11}$$

From (9), (10), and (11) we obtain

$$\overline{\lim_{n\to\infty}} T_a^b(f_n - f_0) \leqq 2(1 + M^2)^{1/2} K(B_0) + 2\, \alpha(M).$$

This relation is independent of S; $\alpha(M)$ can be made arbitrarily small by taking M sufficiently large and, M having been fixed, the first term on the right can be made arbitrarily small by a suitable choice of S. Consequently we have $\lim_{n\to\infty} T_a^b(f_n - f_0) = 0$, and the proof of Theorem 4 is complete.

THEOREM 5. *The relations* $f_n(x) \to f_0(x)$ *on* (a, b) *and* $T_a^b(f_n - f_0) \to 0$, *with* $f_0(x)$ *of bounded variation, imply* $f_n(x) - l \to f_0(x)$.

For by (1) and (4) we have

$$L_a^b(f_0) \leqq \varliminf_{n\to\infty} L_a^b(f_n) \leqq \varlimsup_{n\to\infty} L_a^b(f_n) \leqq L_a^b(f_0) + \varlimsup_{n\to\infty} T_a^b(f_n - f_0) = L_a^b(f_0).$$

THEOREM 6. *The relations* $f_n(x) - l \to f_0(x)$ *and* $g_n(x) - l \to g_0(x)$ *on* (a, b), *with* $f_0(x)$ *and* $g_0(x)$ *absolutely continuous, imply both*

$$[f_n(x) + g_n(x)] - l \to [f_0(x) + g_0(x)] \text{ and } f_n(x)g_n(x) - l \to f_0(x)g_0(x).$$

From (3), (5), and Theorem 4 we infer

$$T_a^b(f_n + g_n - f_0 - g_0) \leqq T_a^b(f_n - f_0) + T_a^b(g_n - g_0) \to 0,$$

$$T_a^b(f_ng_n - f_0g_0) \leqq M \cdot T_a^b(f_n - f_0) + N \cdot T_a^b(g_n - g_0) \to 0,$$

M being a uniform upper bound for $|g_n|$ and N an upper bound for $|f_0|$. The conclusions then follow from Theorem 5.

BROWN UNIVERSITY.

ABELIAN SUBGROUPS OF ORDER p^m OF THE I-GROUPS OF THE ABELIAN GROUPS OF ORDER p^n AND TYPE 1, 1, 1, $\cdots$

By Henrietta Terry

The group of isomorphisms I of a given group is important in the construction of new groups which contain the given group as an invariant subgroup. However, little is known about groups of isomorphisms in general, and even when the given group is of the simplest type, i.e., abelian, the group of isomorphisms has not been thoroughly studied except in a few extremely special cases.

For instance, the I-group of the cyclic group has been discussed by Burnside[1] and Miller.[2] The I-group of a group of order p^n and type $n - 1, 1$ was studied by Miller[3] and although the general type has been barely touched upon he has proved several general theorems.[4] The I-group of the abelian group H of order p^n and type $1, 1, 1, \cdots$ was considered by Moore,[5] and it is well known that the operators U of this group can be represented as non-singular linear transformations on the exponents of n independent generators of H. Thus a determination of the subgroups of the group of these n-ary linear homogeneous transformations modulo p is equivalent to a determination of the subgroups of the I-group of H. Dickson determined all the subgroups in the case $n = 3$,[6] and the subgroups of order a multiple of p in the case $n = 4$,[7] and all the subgroups of the three highest powers of p for all positive integral values of n.[8]

A necessary and sufficient condition that an operator U in I be of order a power of p is that the characteristic determinant of U be $(-1)^n (\lambda - 1)^n$.[9] The invariant factors of such operators are powers of $(1 - \lambda)$ which determine the canonical form and conjugates of U. There is a one-to-one correspondence between these powers of $(1 - \lambda)$, the conjugate sets of operators in I whose orders are powers of p, and the partitions of n. Hence, we shall designate an operator U of I of order p^m by the degrees of its invariant factors or by the partition of n to which it corresponds, i.e.,

$$n = n_1 + n_2 + n_3 + \cdots + n_\gamma + n_{\gamma+1} + \cdots + n_\delta,$$

where the terms are ordered so that $n_i \geqq n_{i+1}$, $n_\gamma > 1$, $n_{\gamma+i} = 1$, if $\gamma < \delta$.

Received by the Editors of the American Journal of Mathematics, September 15, 1934, accepted by them, and later transferred to this journal.

[1] W. Burnside, *Theory of Groups of a Finite Order*, 1897, pp. 239–242.

[2] G. A. Miller, Transactions of the Amer. Math. Soc., vol. 4 (1903), pp. 152–160.

[3] G. A. Miller, Transactions of the Amer. Math. Soc., vol. 2 (1901), pp. 259–264.

[4] G. A. Miller, Annals of Mathematics, (2), vol. 3 (1902), pp. 183–184; Amer. Journ. of Math., vol. 36 (1914), pp. 47–52.

[5] E. H. Moore, Bulletin of the Amer. Math. Soc., vol. 2 (1895), pp. 33–43.

[6] L. E. Dickson, Amer. Journ. of Math., vol. 27 (1905), pp. 189–202.

[7] L. E. Dickson, Amer. Journ. of Math., vol. 28 (1906), pp. 1–16.

[8] L. E. Dickson, Quarterly Journ. of Math., vol. 36 (1904–05).

[9] H. R. Brahana, Proc. of the Nat. Acad. of Sci., vol. 18 (1932), p. 724.

Generators of H can be so chosen that U can be written as a set of partial transformations each on n_i, $i = 1, 2, 3, \cdots, \delta$, generators distinct from those transformed by the other $\delta - 1$ partial transformations.[10] Dickson designates the independent generators which are obtained from one of these partial transformations as a *chain* of length n_i. In the following pages we shall refer to these partial transformations as chains of length n_i in U.

The partitions of n are ordered so that the partition

$$n_1 + n_2 + \cdots + n_\gamma + 1 + 1 + 1 + \cdots$$

is *greater* than the partition

$$n_1' + n_2' + \cdots + n_{\gamma'}' + 1 + 1 + 1 + \cdots,$$

if (1) $n_1 + n_2 + \cdots + n_\gamma \geqq n_1' + n_2' + \cdots + n_{\gamma'}'$, (2) $n_1 \geqq n_1'$ and (3) at least one $n_i > n_i'$. For example, of the two partitions

$$a + b + 4 + 1 + 1 + \cdots$$

and

$$a + b + 2 + 2 + 1 + 1 + \cdots$$

the first is the greater. This ordering of the partitions of n makes it possible to characterize simply a set of abelian subgroups of order p^m of I whose conjugates contain every abelian subgroup of order p^a of I.

The operators in I of the type

$$(1) \qquad \begin{vmatrix} 1 & a_{12} & a_{13} & a_{14} & \cdots & a_{1n} \\ 0 & 1 & a_{23} & a_{24} & \cdots & a_{2n} \\ 0 & 0 & 1 & a_{34} & \cdots & a_{3n} \\ 0 & 0 & 0 & 1 & \cdots & a_{4n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{vmatrix}$$

form a group which is a Sylow subgroup of I. We shall designate (1) by I_p in the following pages. The maximal abelian subgroups of I_p, in which every operator corresponds to a partition of k or less chains of length 2, have been determined.[11]

We proceed to obtain the form and the order of the maximal abelian subgroups of I_p in which the greatest partition to which any operator in the group corresponds is $n_1 + n_2 + \cdots + n_\gamma + 1 + 1 + 1 + \cdots$. The groups are determined throughout by obtaining all of the operators of I_p which are commutative

[10] L. E. Dickson, *Modern Algebraic Theories*, 1926, p. 90.

[11] H. R. Brahana, *On metabelian groups*, Amer. Journ. of Math., vol. 51 (1934), pp. 490–510.

with an element U_1 of the type and which with U_1 do not generate an operator corresponding to a greater partition.

The form chosen for U_1 corresponding to the partition $3 + 1 + 1 + \cdots$ is

$$U_1 = \begin{vmatrix} 1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{vmatrix}.$$

The operators in I_p commutative with U_1 uniquely determine the form of the maximal abelian subgroup except for chains on $S_3, \cdots, S_n$. Obviously U_1 and any operator not commutative with $S_3, \cdots, S_n$ will generate an operator corresponding to a larger partition. Given any partition to which an operator is to correspond a simple form analogous to U_1 can be written.

Operators corresponding to partitions which have k equal terms different from 1 are considered first. The maximal abelian subgroup corresponding to the partition $4 + 4 + 4 + 1 + 1 + 1 + \cdots$ which follows indicates the form for the general case of k chains of length n_1.

Type III_4

$$\begin{vmatrix} 1 & a_{12} & a_{13} & 0 & 0 & 0 & 0 & 0 & 0 & a_{1,10} & a_{1,11} & a_{1,12} & \cdots & a_{1n} \\ 0 & 1 & a_{12} & 0 & 0 & 0 & 0 & 0 & 0 & a_{13} & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & a_{12} & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & a_{45} & a_{46} & 0 & 0 & 0 & a_{4,10} & a_{4,11} & a_{4,12} & \cdots & a_{4n} \\ 0 & 0 & 0 & 0 & 1 & a_{45} & 0 & 0 & 0 & 0 & a_{46} & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & a_{45} & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & a_{78} & a_{79} & a_{7,10} & a_{7,11} & a_{7,12} & \cdots & a_{7,n} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & a_{78} & 0 & 0 & a_{79} & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & a_{78} & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 1 \end{vmatrix}.$$

Since there are $n - 7$ independent a's to be determined in each of the 3 chains the order of the group is $p^{3(n-7)}$. The results and generalizations for k chains of equal length are tabulated below.

Type	Order of Maximal Abelian Subgroup	Generalizations	
I_2	p^{n-1}		
I_3	p^{n-1}	I_{n_1}	p^{n-1}
I_4	p^{n-1}		
II_2	$p^{2(n-2)}$		
II_3	$p^{2(n-3)}$	II_{n_1}	$p^{2(n-n_1)}$
II_4	$p^{2(n-4)}$		
III_2	$p^{3(n-3)}$		
III_3	$p^{3(n-5)}$	III_{n_1}	$p^{3(n-2n_1+1)}$
III_4	$p^{3(n-7)}$		
IV_2	$p^{4(n-4)}$		
IV_3	$p^{4(n-7)}$	IV_{n_1}	$p^{4(n-3n_1+2)}$
IV_4	$p^{4(n-10)}$		
K_2	$p^{k(n-k)}$		$p^{k[n-(k-1)n_1+(k-2)]}$
K_3	$p^{k(n-2k+1)}$	K_{n_1}	$p^{k[n-k(n_1-1)+(n_1-2)]}$
K_4	$p^{k(n-3k+2)}$		$p^{n-\gamma+\delta(\gamma-1)}$

The operators corresponding to partitions whose terms different from 1 are unequal were studied next and the form of the maximal abelian subgroup of I_p which corresponds to the general partition

$$n_1 + n_2 + \cdots + n_\gamma + 1 + 1 + 1 + \cdots$$

determined. This includes the groups corresponding to partitions of n having k equal terms, described in the preceding pages, and a check on the following results is obtained by substituting equal n_i and comparing with the results given in the table.

In the first row of the chain of length n_1 there are $n_i - 1$ zeros for each following chain while the second, third, $\cdots$, and $(n_1 - 1)^{\text{st}}$ rows have only dependent a's. Likewise the first row of the chain of length n_2 has $n_i - 1$ zeros for all following chains and the remaining rows have only dependent a's. The first row of the chain of length n_1 has $n_2 - 1$ zeros due to the chain of length n_2, but the first row of the chain of length n_2 has $n_1 - 1$ zeros to the left of the main diagonal due to the chain of length n_1. Hence the number of independent a's in the first row of the chain of length n_2 is $n_1 - n_2$ less than the number of independent a's in the first row of the chain of length n_1. This can be extended readily to n_i for all i.

It follows that the order of the group is p with the exponent $\alpha_1 + \cdots + \alpha_\gamma$, where $\alpha_j = n - [1 + n_2 - 1 + n_3 - 1 + \cdots + n_\gamma - 1 + (n_1 - n_j)]$. Collecting exponents we have

$$p^{n\gamma-[(\gamma-1)(n_1+n_2+\cdots+n_\gamma)-\gamma(\gamma-2)]},$$

or

$$p^{n-\gamma+\delta(\gamma-1)}.$$

Every subgroup of order p^m, $m < n$, is contained in at least one Sylow subgroup of order p^n, hence the groups determined are maximal abelian subgroups of order p^m of I of H.

Furthermore, the operators in I of order a power of p corresponding to a given partition are conjugate, hence each is in the maximal abelian subgroup corresponding to its partition determined above or one of its conjugates in I. Moreover, every abelian subgroup of order p^m in I is conjugate to an abelian subgroup of I_p. Consequently it is in a conjugate of the maximal abelian subgroup corresponding to the greatest partition to which any operator in the subgroup corresponds. Therefore the abelian subgroups of order p^m in I are completely determined when the abelian subgroups of these maximal abelian subgroups are characterized.

In order to accomplish this task for subgroups of order p^2 it is expedient to consider the groups of order p^{n+2} which contain H as a maximal abelian invariant subgroup. Miller[12] determined all the groups of order p^{n+1}, p being any prime, which contain the abelian group of order p^n and of type 1, 1, 1, $\cdots$.

We proceed to determine the groups G_2 of order p^{n+2} which contain the abelian group H of order p^n and of type 1, 1, 1, $\cdots$ invariantly, generated by H and two independent commutative operators of order p in I corresponding to small partitions of n. In the groups first considered the operator in $\{U_1, U_2\}$ corresponding to the greatest partition has only one chain and, if this operator has a chain of length greater than 2, there is at least one operator, hence a subgroup of order p, in $\{U_1, U_2\}$ corresponding to a lesser partition.

Moreover, two commutative operators of I of given chain length can not generate an operator with a chain of greater length than either, although some operator in the group generated by them may contain more chains than either. From this and the above we have the following

Theorem 1. *If a group of order p^2 is generated by two operators of I with chains of length n_1 and n_2, respectively, on n_1 independent generators of H, $n_1 > n_2$, then $(p - 1)$ of the operators have a chain of length n_2 and the remaining ones, except the identity, a chain of length n_1.*

Using this theorem we determine the number of abstractly distinct groups of order p^{n+2}, of class[13] 3, and with central[14] of order p^{n-2} containing H as a maximal abelian invariant subgroup, extended by 2 commutative operators of order p in I.

In order to have at least one operator with a chain of length 3, we shall define U_1 by

$$U_1^{-1}S_1U_1 = S_1S_2, \qquad U_1^{-1}S_2U_1 = S_2S_3,$$

and let U_2 satisfy the relations

$$U_2^{-1}S_1U_2 = S_1S_k, \qquad U_2^{-1}S_2U_2 = S_2S_3^a.$$

[12] G. A. Miller, Bulletin of Amer. Math. Soc., vol. 8 (1901–02), pp. 391–396.

[13] The class of G_2 is equal to n_1 of U_1.

[14] The order of the central of G_2 is equal to the number of terms in the partition to which U_1 corresponds.

The independent generators of H invariant under U_1 and U_2 can be renamed so that S_k may have the form $S_k = S_2^\alpha S_3^\beta S_4^\gamma$, and by Theorem 1 α can be made zero, hence the most complex form of S_k which we need to consider is $S_k = S_3 S_4^\alpha$ as the exponent of S_3 will be 1 for some power of U_2.

When $\alpha = 0$, we have a group abstractly distinct from the case $\alpha \neq 0$, as in the first case the commutator subgroup K of G_2 is of order p^2, while in the second it is of order p^3. The condition $\alpha = 0$ uniquely determines U_2 and the substitution $S_4' = S_3 S_4^\alpha$ changes an operator with the exponent of S_3 not equal to zero to one in which it is zero. This concludes the consideration of $G_2 = \{H, U_1, U_2\}$, where U_1 and U_2 are of type I_3, or less, with one group for K of order p^3, and one for K of order p^2, each of which has one subgroup of type I_2.

Further application of Theorem 1 shows that for the groups of order p^{n+2} of class 4 and central of order p^{n-3} obtained as before there are four distinct groups as follows:

Order of K	No. of groups	No. of chains of length 2 in U_2
p^4	2	1, 2
p^3	2	1, 2.

And for a group of order p^{n+2} and class n_1 with a central of order p^{n-n_1+1} containing H as a maximal invariant abelian subgroup generated by H, U_1, U_2, U_1 and U_2 being commutative operators of order p and U_2 having one or two chains of length 2, we have the following:

Order of K	No. of groups	No. of chains of length 2 in U_2
p^{n_1}	2	1, 2
p^{n_1-1}	2	1, 2.

At this point we call attention to the kind of operators which appear in the maximal abelian subgroup of type I_{n_1}:

$$\left|\begin{array}{ccccccccc}
1 & a_{12} & a_{13} & a_{14} & \cdots & a_{1,n_1} & a_{1,n_1+1} & \cdots & a_{1,n} \\
0 & 1 & a_{12} & a_{13} & \cdots & a_{1,n_1-1} & 0 & \cdots & 0 \\
0 & 0 & 1 & a_{12} & \cdots & a_{1,n_1-2} & 0 & \cdots & 0 \\
0 & 0 & 0 & 1 & \cdots & a_{1,n_1-3} & 0 & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & 0 & 0 & 0 & \cdots & a_{12} & 0 & \cdots & 0 \\
0 & 0 & 0 & 0 & \cdots & 1 & 0 & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 1
\end{array}\right|.$$

If $a_{12} = a_{13} = \cdots = a_{1,n_1-1} = 0$, the operator is of type I_2.

If $a_{12} = a_{13} = \cdots = a_{1,n_1-2} = 0$, the operator is of type II_2.

If $a_{12} = a_{13} = \cdots = a_{1,n_1-3} = 0$, the operator is of type III_2.

. . . .

If $a_{12} = a_{13} = \cdots = a_{1,n_1-k} = 0$, the operator is of type K_2, where $k \leqq n_1/2$.

If $a_{12} = 0$, $a_{13} \neq 0$, there are two cases:

(1) $n_1 = 2c - 1$ One chain of length c and one chain of length $c - 1$.

(2) $n_1 = 2c$ Two chains of length c.

If $a_{12} = a_{13} = 0$, $a_{14} \neq 0$, there are 3 cases:

(1) $n_1 = 3c - 2$ One chain of length c and two chains of length $c - 1$.

(2) $n_1 = 3c - 1$ Two chains of length c and one chain of length $c - 1$.

(3) $n_1 = 3c$ Three chains of length c.

If $a_{12} = a_{13} = a_{14} = 0$, $a_{15} \neq 0$, there are four cases:

(1) $n_1 = 4c - 3$ One chain of length c and three chains of length $c - 1$.

(2) $n_1 = 4c - 2$ Two chains of length c and two chains of length $c - 1$.

(3) $n_1 = 4c - 1$ Three chains of length c and one chain of length $c - 1$.

(4) $n_1 = 4c$ Four chains of length c.

. . . .

If $a_{12} = a_{13} = \cdots = a_{1j} = 0$, $a_{1,\,j+1} \neq 0$, $j < n_1$, there are j cases:

(1) $n_1 = jc - j + 1$ One chain of length c and $j - 1$ chains of length $c - 1$.

(2) $n_1 = jc - j + 2$ Two chains of length c and $j - 2$ chains of length $c - 1$.

. . . .

(j) $n_1 = jc$ j chains of length c.

When $c = 2$, the chains of length 2, listed above, appear.

The G_2's of class two and central of order p^{n-2} generated by H and two commutative operators U_1 and U_2 of order p in I have been determined.[15] We proceed to consider the G_2's such that the greatest partition to which any operator in $\{U_1, U_2\}$ corresponds is $3 + 2 + 1 + 1 + 1 + \cdots$. Obviously they are of class 3 and have a central of order p^{n-3}. From Theorem 1 we see there is always a subgroup of order p in $\{U_1, U_2\}$ of type II_2 or less and we shall select one of these operators of lesser type for U_2. The commutator subgroup K

[15] H. R. Brahana, *On metabelian groups*, loc. cit. (footnote 11), section 6.

may be of order p^5, p^4, p^3. Each case is examined for the number and type of subgroups of lesser partition in $\{U_1, U_2\}$. Obviously these subgroups differentiate between groups which are not simply isomorphic.

However, in the groups G_2 with K of order p^4 there are 2 groups having no operators of type I and only 1 subgroup of order p of type II_2 in $\{U_1, U_2\}$ which are not simply isomorphic. In one of these the commutator subgroup of $\{U_2, S_1\}$ is identical with the group generated by the commutator of a commutator in $G = \{H, U_1\}$ which is a characteristic subgroup in G, and this relation does not exist in the other, hence these two groups are not simply isomorphic.

We meet this situation again in the case K is of order p^3 and U_2 generates a subgroup of order p of type I_2, the only subgroup corresponding to a partition less than $3 + 2 + 1 + 1 + 1 + \cdots$ in $\{U_1, U_2\}$.

The groups for G_2 of order p^{n+2} and class 3 with a central of order p^{n-3} containing H as a maximal abelian invariant subgroup of order p^n extended by 2 commutative operators of order p in I are as follows:[16]

Order of K	No. of groups	No. of subgroups of type I_3	I_2
p^5	1	0	0
p^4	1	1	0
	1	0	1
	2	0	0
p^3	1	1	1
	1	1	0
	2	0	1
	1	0	0.

In addition to these subgroups just mentioned we have determined all the subgroups of order p^{n+2} in the holomorph of H whose operators are all of order p, whose central is of order p^{n-n_1+1}, whose class is n_1, whose commutator subgroup is of order p^{n_1-1}, whose cross-cut of the central and the commutator subgroup is of order p and which contain subgroups of order p^{n+1} of classes n_1 and 2 only, the commutator subgroup of the latter being of order p^2 at most.

The results obtained indicate the large number and variety of subgroups of order p^2 in I_p as well as the number of abstractly distinct groups of order p^{n+2} which contain the abelian group of order p^n and type 1, 1, 1, $\cdots$ invariantly. The complete characterization of the groups of order p^2 in I is a long and difficult task but the methods which we have developed and applied in this paper are apparently sufficient to complete it.

UNIVERSITY OF ILLINOIS.

[16] The groups on lines 4 and 7 of the table are differentiated by the condition that in one of them the commutator subgroup of $\{U_2, S_1\}$ is identical with the subgroup generated by the commutator of a commutator in $G = \{H, U_1\}$ which is a characteristic subgroup in G.

GENERALIZED PERFECT SETS

By G. T. Whyburn

1. In an earlier paper[1] the author has defined generalized derived aggregates or K-derivatives for subsets A of a metric space C with respect to an arbitrary class K of closed sets in C. Under this definition, by the K-derivative $K(A)$ of A is meant the set of all points x such that every neighborhood of x contains a subset of A which is not contained in any K-set. The operation of taking K-derivatives may be iterated, giving successive K-derivatives denoted as follows: $A = A_K^0, K(A) = A_K^1, \cdots , K(A_K^{n-1}) = A_K^n, \cdots$, and in general $A_K^\alpha = K(A_K^{\alpha-1})$ or $= \prod_{\beta<\alpha} K(A_K^\beta)$ according as α is an isolated or a limit ordinal.

This suggests the following extension of the notion of a perfect set, to which the present paper is devoted. A set of points A will be said to be *K-perfect* provided it is equal to its own K-derivative, i.e., $K(A) = A$. It results at once that, for all classes K, every K-perfect set is closed. In case K is the class of all single points, then the K-perfect sets reduce to the ordinary perfect sets.

Other examples are: (1) the Sierpinski triangle curve[2] is perfect with respect to the class of all simple closed curves, the class of all simple arcs, or the class of all dendrites; but it is not perfect relative to the class of all regular curves; (2) the set E consisting of the curve $y = \sin 1/x$ together with the interval $I = (-1, 1)$ of the y-axis is perfect with respect to the class of all simple closed curves in E, (a vacuous class); but relative to the class of simple arcs in E, this set is not perfect, since its first arc-derivative reduces to I and its second arc-derivative vanishes; (3) in the set F consisting of the curve $y^2 = x^2 \sin^2 1/x$ together with the origin Q, the point Q itself is perfect with respect to the class K of all simple closed curves contained in F; in fact, in this case $F_K^1 = Q$, $F_K^2 = K(Q) = Q$, and so on.

In what follows we shall suppose our space to be separable and metric, and K will denote an arbitrary class of closed sets in this space.

2. Theorem. *Any closed set A is the sum of a K-perfect set and a countable number of sets each of which is the intersection of A with a K-set.*

Proof. Let B be the first ordinal such that $A_K^\beta = A_K^{\beta+1}$. Since the space is separable, β is of the first or second class. Then A_K^β is K-perfect and we have

$$A = A_K^\beta + \sum_{0 \leq \alpha < \beta} [A_K^\alpha - A_K^{\alpha+1}]. \tag{i}$$

Now for any α, each point x of $A_K^\alpha - A_K^{\alpha+1}$ is contained in some neighborhood Q^α such that $A_K^\alpha \cdot Q^\alpha$ is contained in some K-set. Whence, by the Lindelöf

Received February 13, 1935.

[1] See American Journal of Mathematics, vol. 54 (1932), pp. 169–175.

[2] See Comptes Rendus, vol. 162, p. 629.

theorem, for each α the set $A_K^\alpha - A_K^{\alpha+1}$ can be covered by a countable collection of neighborhoods $[Q_i^\alpha]$ such that for each i, $A_K^\alpha \cdot Q_i^\alpha$ is contained in some K-set K_i^α. Thus

$$\text{(ii)} \qquad \sum_{0 \leq \alpha < \beta} [A_K^\alpha - A_K^{\alpha+1}] \subset \sum_{0 \leq \alpha < \beta} \sum_{i=1}^{\infty} K_i ,$$

and since β is of the first or second class, the number of sets $[K_i^\alpha]$ on the right is countable.

Now (i) and (ii) together give

$$\text{(iii)} \qquad A = A_K^\beta + \sum_{0 \leq \alpha < \beta} \sum_{i=1}^{\infty} A \cdot K_i^\alpha ,$$

which proves our theorem.

Notes. (1) In case K is the class of single points, the theorem just established reduces to the classical result that any closed set is the sum of a perfect set and a countable set.

(2) If A is the whole space, or if all the K-sets are contained in A, or if every closed subset of a K-set is likewise a K-set, then all sets of the form $A \cdot K$ are K-sets and our theorem takes the following form: *The set A is the sum of a K-perfect set and a countable number of K-sets.*

(3) Since by definition every K-set must be nowhere dense in any given K-perfect set, and since by a well known theorem no self-compact set is the sum of a countable number of nowhere dense sets, it follows that *no compact K-perfect set is contained in the sum of a countable number of K-sets.* This result together with our theorem above gives the following proposition:

In order that a self-compact set A should be contained in the sum of a countable number of K-sets, it is necessary and sufficient that A contain no K-perfect subset.

3. It is possible for two quite distinct classes of sets K to yield the same K-perfect sets. For example, the class of all dendrites and the class of all sets which locally are dendrites in a given space yield identical K-perfect sets. Also, in general, if a class K_1 is contained in a class K_2, then any K_2-perfect set must be K_1-perfect. Similarly, any set which is perfect relative to the class of all arcs in a given space must also be perfect with respect to the class of all simple closed curves in this space. In studying relationships of this sort we are led to the following

THEOREM. *Given two classes K_1 and K_2 of closed sets. If no K_1-set contains a K_2-perfect set, then every K_2-perfect set is likewise K_1-perfect.*

For let A be any K_2-perfect set. Then if A were not K_1-perfect, there would exist a point x of A and a neighborhood U of x such that $A \cdot U$ is contained in some K_1-set B. But this is impossible, since B would contain $\overline{A \cdot U}$ and it is easily seen that $\overline{A \cdot U}$ is K_2-perfect.

Notes. (1) In case the space is compact, the theorem just given takes the following equivalent form: *If each K_1-set is contained in the sum of a countable*

number of K_2-sets, then every K_2-perfect set is also K_1-perfect. Thus in particular the class of all arcs and the class of all arc-sums (i.e., continua which are the sum of a countable number of arcs) in a compact space would define the same K-perfect sets.

(2) We have seen from example (3) in §1 that for some classes K, the K-perfect sets may not be perfect in the ordinary sense. However, the theorem of this section readily yields conditions under which this is the case. For, taking K_1 as the class of all single points in a space C and K_2 as an arbitrary class K of closed sets, our theorem tells us that in order for every K-perfect set to be perfect in the ordinary sense it suffices that every point of the set be contained in some K-set.

4. In a recent article[3] Hurewicz has suggested the following method of taking generalized derived aggregates of a metric space X. Let $f(x)$ be a single valued transformation defined on X and mapping X into a space Y. Then f is said to be stationary at the point x with respect to a subset M of X provided that all points of M in some neighborhood of x map into a single point under f. Denote by $X_1(f)$ the set of all points of X where f is not stationary relative to X, by $X_2(f)$ the set of all points of $X_1(f)$ where f is not stationary relative to $X_1(f)$, and so on with the usual iteration and product taking. In this way we obtain a generalized α^{th} derivative $X_\alpha(f)$ of X with respect to f.

In this concluding section it will be shown that these generalized derivatives proposed by Hurewicz are equivalent to the K-derivatives previously treated by the author in the article referred to above in §1 in the sense stated in the following theorem.

THEOREM. *For any class K of closed sets in a metric space X there exists a transformation $f(x)$ defined on X such that $X_\alpha(f) = X_K^\alpha$ for all α; and conversely, for every transformation $f(x)$ defined on X there exists a class K of closed sets such that $X_\alpha(f) = X_K^\alpha$ for all α.*

Proof. To prove the first part we have only to define $f(x)$ on X as follows:

$$f(x) = \alpha, \quad \text{for} \quad x \epsilon [X_K^\alpha - X_K^{\alpha+1}];$$

$$f(x) = \beta, \quad \text{for} \quad x \epsilon X_K^\beta,$$

where β is the first ordinal such that $X_K^\beta = X_K^{\beta+1}$. Then since

$$X = X_K^\beta + (X_K^0 - X_K^1) + \cdots + (X_K^\alpha - X_K^{\alpha+1}) + \cdots \qquad (\alpha < \beta),$$

and no two of the sets on the right can intersect, $f(x)$ is defined and single valued on X and maps X into the set of all ordinal numbers α, $(0 \leqq \alpha \leqq \beta)$. Now if α is any isolated ordinal, then since $f(x) = \alpha - 1$ on $X_K^{\alpha-1} - X_K^\alpha$ and $f(x) \geqq \alpha$ on X_K^α, it follows that relative to $X_K^{\alpha-1}$, f is stationary on exactly the set $X_K^{\alpha-1} - X_K^\alpha$ and hence is nowhere stationary on exactly the set X_K^α. This gives $X_K^\alpha = X_\alpha(f)$

[3] See Fundamenta Mathematicae, vol. 23 (1934), footnote to p. 54.

for all such α's, and a simple application of transfinite induction yields this same relation for all ordinals α.

Now conversely, if we are given $f(x)$ defined on X, then if we let K be the class of sets $[f^{-1}(y)]$ for all $y \epsilon f(X)$, we will have $X_\alpha(f) = X_K^\alpha$ for all α's. For if α is any isolated ordinal, and x is any point of $X_{\alpha-1}(f)$ where f is stationary relative to $X_{\alpha-1}(f)$, there exists a neighborhood V of X such that $V \cdot X_{\alpha-1}(f)$ is contained in a K-set $f^{-1}(y)$, whereas if f is not stationary at x relative to $X_{\alpha-1}(f)$, then no such neighborhood of x exists. This gives $X_K^\alpha = X_\alpha(f)$ for all isolated α's, and again the principle of transfinite induction yields this relation for all α's.

THE UNIVERSITY OF VIRGINIA.

RIEMANNIAN MANIFOLDS IN THE LARGE

By Sumner Byron Myers

1. Introduction. The general problem to be considered here is that of determining relations between the local properties of an analytic Riemannian manifold and the topological properties of the manifold in the large. In particular, we are interested in determining topological properties from a knowledge of local properties in the neighborhood of just one point, and conversely, in determining possibilities of metrisation of a given topological manifold by means of local analytic Riemannian geometries.

By an n-dimensional analytic Riemannian manifold in the large will be meant a complete manifold in a sense to be defined later, equivalent to the "normal" Riemannian space of Cartan[1] and a generalization to n dimensions of the "complete surface" of Hopf and Rinow.[2]

The results obtained here are in most cases generalizations of theorems given in the two-dimensional case by Hopf and Rinow. A complete summary of known results on the general problems stated above, as well as a statement of some specific unsolved problems can be found in a paper by Hopf entitled *Differentialgeometrie und topologische Gestalt*, Jahresbericht der Deutschen Mathematiker-Vereinigung, vol. 41 (1932), pp. 209–229. Some of the problems proposed there (p. 222, lines 14–22, and p. 224, lines 25–29) are solved in the present paper.

In §2 we define *complete manifolds*. §3 contains a proof that a complete manifold whose Riemannian curvature at every point and with respect to every plane direction is greater than a positive constant e is compact and has a diameter less than $\pi/e^{\frac{1}{2}}$. In §4 is proved the fundamental uniqueness theorem—a given n-dimensional Riemannian element E can be continued to (i.e., contained in) at most one complete, simply connected n-dimensional manifold. In §5 we set up certain coördinate systems in E and give necessary analyticity conditions that E may be continued to a complete manifold. We also show how to determine from a certain coördinate system in the element E about a point A the points conjugate to A. In §6 we show that under certain analyticity conditions an element E about a point A without conjugate point can be continued to a complete manifold homeomorphic to n-dimensional space S_n, from which follows that if a complete simply connected n-dimensional manifold contains a point without conjugate point, it is homeomorphic to S_n. Finally in §7 we

Received by the Editors of the Annals of Mathematics, February 27, 1934, accepted by them, and later transferred to this journal; presented to the American Mathematical Society December, 1933.

[1] E. Cartan, *Leçons sur la géométrie des espaces de Riemann*, Paris, 1928.

[2] H. Hopf and W. Rinow, *Über den Begriff der vollständigen differentialgeometrischen Fläche*, Commentarii Mathematici Helvetici, vol. 3 (1931).

prove that if the universal covering manifold of a complete manifold is not compact, through every point A passes at least one geodesic ray without conjugate point to A.

2. **Complete analytic n-dimensional Riemannian manifolds.**[3] We are concerned here with *topological n-dimensional manifolds* ($n \geqq 2$). These are connected separable Hausdorff spaces, in which each neighborhood of the denumerable set of neighborhoods used to define the space is homeomorphic to the interior of the $(n - 1)$-dimensional unit sphere. Only intrinsic properties of these manifolds will be used, so that no imbedding in euclidean spaces is necessary. The words *closed* and *open* as applied to these manifolds are synonymous with compact and not compact in the usual sense.

If each of the neighborhoods is provided with a coördinate system so that in the region common to two intersecting neighborhoods one set of coördinates can be obtained from the other by means of an analytic transformation with non-vanishing jacobian, the manifold is called *analytic*. Any coördinate system which is obtainable from one of the above by means of an analytic transformation with non-vanishing jacobian is said to be *admissible*.

Now suppose that the manifold is metrisable in the following manner. To each neighborhood with admissible coördinate system $(x) = (x_1, \cdots, x_n)$, we assign a real analytic positive definite symmetric quadratic differential form

$$g_{\alpha\beta} dx_\alpha dx_\beta \qquad (\alpha, \beta = 1, \cdots, n).$$

The functions appearing in these quadratic forms are to have the property that if (x) and $(\bar{x})$ are admissible coördinate systems respectively in two intersecting neighborhoods, and $g_{\alpha\beta}(x)dx_\alpha dx_\beta$ and $\bar{g}_{\alpha\beta}(\bar{x})d\bar{x}_\alpha d\bar{x}_\beta$ are the quadratic forms respectively assigned to them, then in the intersection of the two neighborhoods,

$$g_{\alpha\beta}(x)dx_\alpha dx_\beta = \bar{g}_{\alpha\beta}(\bar{x})d\bar{x}_\alpha d\bar{x}_\beta$$

under the transformation between (x) and $(\bar{x})$. We define arc length along a curve as the integral of the square root of the quadratic differential form. We now have an *analytic n-dimensional Riemannian manifold M*.

If we define the distance between two points as the lower limit of the lengths of rectifiable curves joining the two points, the manifold M satisfies the axioms for a metric space. Furthermore, the idea of neighborhood resulting from this notion of distance coincides with the original topological notion of neighborhood.

In treating such manifolds in the large we restrict ourselves to manifolds which are *complete*. Our definition of completeness is identical with that given by Hopf and Rinow in the two-dimensional case, and is as follows.

An analytic n-dimensional Riemannian manifold M is said to be *complete* if it satisfies any one of the following four equivalent postulates:

(1) Every geodesic ray can be continued to infinite length on M.

[3] Statements made in §2, requiring proof, for which no proof is given, are formal generalizations of corresponding 2-dimensional theorems and proofs of Hopf and Rinow, loc. cit.

(2) Every divergent[4] line is infinitely long.

(3) Every fundamental sequence[5] of points on M converges.

(4) Every bounded set of points in M has a limit point in M.

Postulates (1) and (2) have been used in the treatment of analytic Riemannian manifolds of constant curvature.[6] Postulate (3) makes M a complete metric space in the sense of Hausdorff,[7] while postulate (4) (the Weierstrass-Bolzano theorem) is more restrictive than (3) for general metric spaces. However, for analytic Riemannian manifolds the four postulates are equivalent.[8]

Throughout the rest of this paper the word *manifold* will be used to denote an analytic Riemannian manifold.

An n-dimensional manifold M which is complete in the sense just defined is an *entire* or *non-continuable* manifold. By this we mean that M is not a proper subset of another n-dimensional manifold $\bar{M}'$. However, the class of complete manifolds is smaller than the class of non-continuable manifolds. An example of an incomplete but non-continuable n-dimensional manifold is the universal covering manifold $\bar{M}$ of the manifold M obtained by removing an $(n - 2)$-dimensional euclidean space from n-dimensional euclidean space. It is understood that the neighborhood of each point of $\bar{M}$ is to be provided with the euclidean geometry of the neighborhood of M which it covers. Similar examples are the universal covering manifolds respectively of the manifold obtained by removing an $(n - 2)$-dimensional hyperbolic space from n-dimensional hyperbolic space, and of the manifold obtained by removing an $(n - 2)$-sphere from the n-sphere.

Nevertheless, the class of complete manifolds is not too small for our purposes. It contains all closed manifolds, and also all open manifolds which can be regularly imbedded in a euclidean space and are closed in that space. Furthermore, a number of fundamental theorems can be proved for complete manifolds but not for the more general class of non-continuable manifolds.

A first such theorem is the following:

THEOREM 1. *Every two points on a complete manifold can be joined by a curve of shortest length, and this curve is a geodesic.*

The examples given above illustrate the fact that this theorem does not hold if the hypothesis of completeness is replaced by that of non-continuability.

[4] By a *divergent line* on a manifold M is meant the single-valued continuous image of a straight line ray if to every divergent sequence of points on the ray corresponds a divergent sequence of points on M.

[5] A *fundamental sequence of points* x_i in a metric space is a sequence for which the distance $\rho(x_j, x_{j+1})$ satisfies the Cauchy criterion.

[6] See H. Hopf, *Zum Clifford-Kleinschen Raumproblem*, Math. Annalen, vol. 95 (1926), pp. 313–315.

Also see P. Koebe, *Riemannsche Mannigfaltigkeiten und nichteuklidische Raumformen*, Sitzungsber. Preuss. Akad. d. Wiss., Phys.-math. Klasse, Berlin, (1927), pp. 184–185 and (1928), pp. 349–350.

[7] See F. Hausdorff, *Mengenlehre*, p. 315.

[8] The proof of this fact, and the proof of Theorem 1 which follows later, are formal

3. **Relations between curvature and topological properties.** The following are well known results on space-forms (manifolds of constant curvature).

(A) *For every n and every K there exists a unique*[9] *complete simply connected n-dimensional space-form with curvature K. According as $K > 0$, $= 0$, or < 0 this manifold is a spherical, euclidean, or hyperbolic space.*

(B) *A complete n-dimensional space-form with positive curvature is closed.*[10]

We give now a statement and proof of a theorem which is a generalization of (B) to manifolds of non-constant curvature and at the same time a generalization to n dimensions of a 2-dimensional theorem of Hopf and Rinow (loc. cit., p. 224, Satz V). This result includes a generalization of the classical Bonnet theorem on the diameter of an ovaloid.[11]

THEOREM 2. *A complete n-dimensional manifold M whose curvature is greater than a positive constant e at every point and with respect to every plane direction is closed and has a diameter less than $\pi/e^{\frac{1}{2}}$.*

Consider any geodesic arc g on M. We can set up coördinates $(x) = (x_1, \cdots, x_n)$ in the neighborhood of g such that the functions $g_{\alpha\beta}(x)$ $(\alpha, \beta = 1, \cdots, n)$ appearing in the fundamental quadratic form satisfy the following conditions along g:

$$(3.1) \qquad \begin{aligned} g_{\alpha\beta} &= \delta_{\alpha\beta} \\ \frac{\partial g_{\alpha\beta}}{\partial x_\gamma} &= 0 \end{aligned} \qquad (\alpha, \beta, \gamma = 1, \cdots, n)$$

while the coördinates $x_1, \cdots, x_{n-1}$ are constant along g and x_n is the arc length s measured from any point A on g.[12]

The points on g conjugate to A are given by the zeros of a determinant whose columns are $n - 1$ linearly independent sets of solutions $\eta_1(s), \cdots, \eta_{n-1}(s)$ of the equations

$$(3.2) \qquad \eta_i'' + R_{nk,ni}(s)\eta_k = 0 \qquad (i, k = 1, \cdots, n-1)$$

vanishing at $s = 0$ but not identically zero. The functions $R_{\alpha\beta,\gamma\delta}(s)$ $(\alpha, \beta, \gamma, \delta = 1, \cdots, n)$ are Riemann symbols of the first kind in the coördinates (x) taken along g.

generalizations of the 2-dimensional proofs given by Hopf and Rinow, loc. cit., pp. 215–221. Fundamental in the proof is the fact that every point of a Riemannian manifold has a neighborhood in which every two points can be joined by a unique geodesic arc of shortest length, which has been proved for n dimensions by J. H. C. Whitehead, Quarterly Journal of Mathematics, vol. 3 (1932), pp. 33–42.

[9] That is, unique except for isometric manifolds.

[10] See H. Hopf, Math. Annalen, vol. 95 (1926), pp. 313–339.

[11] A generalization of the Bonnet theorem to n dimensions has previously been given by I. J. Schoenberg, Annals of Math., vol. 33 (1932), pp. 485–495. In the present paper the hypothesis of completeness enables us to draw the topological conclusion that the manifold is closed.

[12] See T. Levi-Civitá, Math. Annalen, vol. 97 (1927), pp. 291–320.

The curvature of M at a point P on g with respect to the plane direction defined by the unit vector $(0, 0, \cdots, 1)$ tangent to g at P and any unit vector $(u_1, \cdots, u_{n-1}, 0)$ orthogonal to g at P is given by

$$R_{nk,ni}u_iu_k \qquad (i, k = 1, \cdots, n-1). \tag{3.3}$$

Then according to hypothesis, along g

$$R_{nk,ni}u_iu_k > e \qquad (i, k = 1, \cdots, n-1) \tag{3.4}$$

for all numbers $(u_1, \cdots, u_{n-1})$ satisfying the relation

$$\sum_{i=1}^{n-1} u_i^2 = 1. \tag{3.5}$$

We may write the inequality (3.4) as follows:

$$R_{nk,ni}u_iu_k > e\sum_{i=1}^{n-1} u_i^2 \qquad (i, k = 1, \cdots, n-1). \tag{3.6}$$

Let us compare the two sets of equations

$$\eta_i'' + R_{nk,ni}\eta_k = 0 \qquad (i, k = 1, \cdots, n-1) \tag{3.7}$$

and

$$\eta_i'' + e\eta_i = 0 \qquad (i = 1, \cdots, n-1). \tag{3.8}$$

According to the generalization of the Sturm comparison theorem given by Morse,[13] we can deduce from (3.6) that the first conjugate point of A on g is at a distance less than $\pi/e^{\frac{1}{2}}$ from A. Thus any geodesic arc on M of length $\pi/e^{\frac{1}{2}}$ or greater contains a point conjugate to its initial point, and is not the shortest arc joining its end points.

Hence the geodesic of shortest length which, according to Theorem 1, exists between any two points of the complete manifold M, is shorter than $\pi/e^{\frac{1}{2}}$. The diameter of M is thus less than $\pi/e^{\frac{1}{2}}$, and M is a bounded manifold. A use of the fourth completeness postulate enables us to conclude that the boundedness of M implies its compactness, that is, M is closed.

4. **The uniqueness theorem.** By an n-dimensional Riemannian element E we shall mean a point and its neighborhood homeomorphic to the interior of the $(n-1)$-dimensional sphere, and provided with an analytic Riemannian geometry, that is, with a coördinate system and an analytic, positive definite, symmetric Riemannian quadratic form. We are concerned with determining the topological properties of any complete manifold to which a given element may be continued.

Among all the manifolds the simply connected ones play an important rôle. For if we have an arbitrary manifold M containing an element E, we can obtain

[13] Morse, Math. Annalen, vol. 103 (1930), p. 66.

a simply connected manifold also containing the element E by providing each neighborhood of the universal covering manifold $\bar{M}$ of M with the Riemannian geometry of the neighborhood of M which it covers. If M is complete, $\bar{M}$ will also be complete.

The following theorem, a generalization of Theorem (A) on manifolds of constant curvature, is fundamental.[14]

THEOREM 3. *Every n-dimensional Riemannian element E can be continued to at most one complete, simply connected n-dimensional manifold, that is, if two such continuations of E exist, they are isometric.*

Suppose that we have two complete simply connected continuations M and M' of the element E. Then the neighborhood N of a certain point A on M will be isometric to the neighborhood N' of a point A' on M'. Any admissible coördinate system in N and its quadratic differential form can be used in N' by giving points in N' the same coördinates as their correspondents in N.

Let g be a geodesic arc issuing from A, and g' the corresponding geodesic arc issuing from A'. We can set up the coördinates (x) of §3 in the neighborhood of g on M. In the neighborhood of g' on M' we can set up a similar coördinate system (x), with the property that $(x') = (x)$ for points in N and N' corresponding under the given isometry.

We will use $R_{\alpha\beta,\gamma\delta}(x)$ to denote the Riemann symbols of the first kind in the coördinates (x). Then $R_{nk,ni}(0, \cdots, 0, s)u_iu_k$ $(i, k = 1, \cdots, n-1)$ is the curvature of M with respect to the plane direction defined by the unit vector tangent to g at P and the unit vector $(u_1, \cdots, u_{n-1}, 0)$ orthogonal to g at P; that is, the curvature at P of the geodesic surface S determined by the direction of g at P and the direction $(u_1, \cdots, u_{n-1}, 0)$. Now in the coördinates (x) the Christoffel symbols are all zero along g, and hence if we move the vector $(u_1, \cdots, u_{n-1}, 0)$ along g parallel to itself with respect to g and M, the components $u_1, \cdots, u_{n-1}, 0$ remain constant. But according to the theorem of Severi we get the same result if we move the vector $(u_1, \cdots, u_{n-1}, 0)$ along g parallel to itself with respect to g and S. Thus the same geodesic surface S is determined by the direction of g at each of its points and the direction $(u_1, \cdots, u_{n-1}, 0)$. Hence $R_{nk,ni}u_iu_k$ is the curvature of the surface S along g.

We conclude that $R_{nk,ni}(s)u_iu_k$ is an analytic function of s for all values of $s > 0$[15] and for all $(u_1, \cdots, u_{n-1})$. Therefore $R_{nk,ni}$ is an analytic function of s for all positive values of s. Similarly, on the manifold M' the Riemann symbols $R'_{nk,ni}$ in the coördinates (x') are analytic functions of s along g'.

But along the arcs of g and g' lying in N and N' respectively $R_{nk,ni} = R'_{nk,ni}$. Hence $R_{nk,ni} = R'_{nk,ni}$ for all values of $s > 0$.

Now the functions $R_{nk,ni}(s)$ and $R'_{nk,ni}(s)$ determine the points on g and g'

[14] For $n = 2$, this theorem and the following theorems have been proved by W. Rinow, *Über Zusammenhänge zwischen der Differentialgeometrie im Grossen und im Kleinen*, Math. Zeit., vol. 35 (1932), pp. 512–528. J. H. C. Whitehead has given a generalization of (A) in a different direction to locally homogeneous spaces with a Lie pseudo-group. See Annals of Math., vol. 33 (1932), pp. 681–687.

[15] Since M is complete, the geodesic arc g can be extended to infinite length.

conjugate to A and A' respectively, through equations (3.2). Hence the points on g conjugate to A are at the same distance from A as the points on g' conjugate to A' are from A'.

Let AB be a smooth segment of g; that is, a segment without multiple points and without a point conjugate to A. The corresponding segment $A'B'$ of g' will, by the result of the preceding paragraph, contain no point conjugate to A'. Then there exists an $(n - 1)$-parameter family of geodesics on M through A

$$x_\alpha = x_\alpha(y_1, \cdots, y_{n-1}, s) \tag{4.1}$$

containing g for $(y_1, \cdots, y_{n-1}) = (0, \cdots, 0)$ and forming a field F[16] in the neighborhood of AB. If $A'B'$ has no multiple points, it can be imbedded in a similar field F' by means of the corresponding field of geodesics through A'. The parameters $y'_1, \cdots, y'_{n-1}$ in this latter family can be taken so that geodesics through A and A' corresponding under the given isometry between the neighborhoods N and N' of A and A' respectively are determined by equal parameters $(y'_1, \cdots, y'_{n-1}) = (y_1, \cdots, y_{n-1})$.

In the field F we can use as coördinates of a point P n numbers $(y_1, \cdots, y_n)$ the first $n - 1$ of which are equal to the parameters in (4.1) determining the geodesic AP of the family (4.1) and the last of which is equal to the length of the geodesic arc AP. In F' we can choose similar coördinates $(y'_1, \cdots, y'_n)$. Points in $F \cdot N$ and $F' \cdot N'$ which correspond under the given isometry will have equal coördinates, and hence the Riemannian quadratic forms in the variables (y) and (y') will be the same in $F \cdot N$ and $F' \cdot N'$, therefore the same throughout F and F'. If we make points in F and F' with the same coördinates correspond, the correspondence will be an isometry between F and F', and will be a continuation of the given isometry between N and N'.

If $A'B'$ has multiple points, the family of geodesics through A' forms a field im kleinen. The variables (y') can be used as coördinates im kleinen, and by making points with equal coördinates correspond, we obtain a single-valued, im kleinen isometric map of the field F on the multiple-leaved field around $A'B'$.

Thus the isometric map of N and N' can be continued along any smooth geodesic arc issuing from A. Furthermore, if P is any point on M which can be reached in this manner, the isometry can be continued along any smooth geodesic arc issuing from P. Since any geodesic arc, or any broken geodesic arc, is composed of a finite number of smooth geodesic arcs, the isometry can be continued from A along any such arc.

Any point Q on M can be connected to A by a geodesic arc. Hence the neighborhood of any point Q of M can be mapped isometrically on a neighborhood of a point Q' of M'.

We must now show how to continue the isometry along an arbitrary arc AP issuing from A. This is done as in the 2-dimensional case[17] by constructing a closely approximating broken geodesic with corners on AP.

[16] An improper field. The field breaks down at A.

[17] See Rinow, loc. cit., pp. 516–518.

Finally, an application of the Monodromiesatz shows that the im kleinen isometric map of M on M' is single-valued, and since the rôles of M and M' can be reversed, the map is one-to-one. Thus we have an isometry between M and M'.

COROLLARY 1. *Let M' be a complete n-dimensional manifold, and M an arbitrary simply-connected n-dimensional manifold. If M' and M are continuations of the same Riemannian element E, then M can be mapped in a single-valued and im kleinen isometric manner on a portion of M'.*

The proof is included in the proof of the theorem.

COROLLARY 2. *If two complete n-dimensional manifolds are continuations of the same Riemannian element E, then their universal covering manifolds are homeomorphic.*

5. Conditions that an element can be continued to a complete manifold.

Not every analytic Riemannian element can be continued to a complete manifold. Some necessary conditions are given in the following theorems.

THEOREM 4. *Given an n-dimensional Riemannian element E around a point A. Set up the geodesic coördinates (x) of §3 along any geodesic arc issuing from A. Then if E can be continued to a complete n-dimensional manifold, the Riemann symbols $R_{nk,ni}$ in the coördinates (x) taken along g must be analytic functions of the arc length $x_n = s$ for all positive values of s.*

This has already been shown in the proof of Theorem 4.

THEOREM 5. *Given an n-dimensional Riemannian element E around a point A. Set up the coördinates (y) of § 4 along the neighborhood of any geodesic arc g issuing from A, thus obtaining a fundamental quadratic form of the type $a_{\alpha\beta}dy_\alpha dy_\beta$ $(\alpha, \beta = 1, \cdots, n)$ in which $a_{nn} = 1$, $a_{in} = 0$ $(i = 1, \cdots, n-1)$. Then if E can be continued to a complete n-dimensional manifold, each function a_{ij} $(i, j = 1, \cdots, n-1)$ taken along g must be an analytic function of the arc length $y_n = s$ for all positive values of s.*

The transformation from the coördinate system (x) in the neighborhood of g to the coördinate system (y) is given by the equations (4.1) of the $(n-1)$-parameter family of geodesics through A. Each column of the determinant $|\partial x_k/\partial y_j|$ taken along g represents a set of variations of the family of geodesics (4.1) along g, and hence is a solution of equations (3.2), since the latter are the equations of variation of the geodesics.

According to Theorem 4, $R_{nk,ni}$ is an analytic function of s for all $s > 0$. Hence solutions of (3.2) are analytic functions of s for all $s > 0$. In particular, along g $\partial x_k/\partial y_j$ $(j, k = 1, \cdots, n-1)$ is an analytic function of s for all $s > 0$.

But along g

$$a_{ij} = \frac{\partial x_k}{\partial y_i}\frac{\partial x_k}{\partial y_j} \qquad (i, j, k = 1, \cdots, n-1). \tag{5.15}$$

Therefore along g a_{ij} is an analytic function of s for all $s > 0$. This completes the proof.

As in the case of polar coördinates on two-dimensional manifolds,[18] it can be shown that the functions $a_{11}, \cdots, a_{n-1,n-1}$ vanish at A. It follows from (5.15) that $\partial x_k/\partial y_i$ $(i, k = 1, \cdots, n-1)$ vanish at A. Hence $|\partial x_k/\partial y_i|$ is a determinant whose zeros $s \neq 0$ define the points on g conjugate to A. But by means of (5.15) we see that along g

$$(5.16) \qquad |a_{ij}| = \left|\frac{\partial x_k}{\partial y_j}\right|^2.$$

We have, then, the following theorem:

THEOREM 6. *The points on g conjugate to A are given by the zeros $s \neq 0$ of the determinant $|a_{ij}|$ taken along g.*

THEOREM 7. *Given an n-dimensional element E around a point A. Set up coördinates (y) as in Theorem 5 along an arbitrary geodesic g through A. Then if E can be continued to a complete n-dimensional manifold M, a_{ij} will be an analytic function of the n variables $(y_1, \cdots, y_n)$ for $0 < y_n < K$, where K is the first value of y_n for which $|a_{ij}|$ vanishes on g, and for $(y_1, \cdots, y_{n-1})$ in the neighborhood of their values on g.*

According to Theorem 6, $y_n = K$ gives the first point on g conjugate to A. Hence the geodesics through A form a field in M in the neighborhood of g for $0 < y_n < K$. We can use the coördinates (y) throughout this field, and the functions a_{ij} will be analytic throughout the field.

COROLLARY. *If $|a_{ij}|$ has no zeros on g, then if E can be continued to a complete manifold, a_{ij} is an analytic function of $(y_1, \cdots, y_{n-1})$ in the neighborhood of g for $0 < y_n < \infty$.*

6. **Conjugate points and manifolds in the large.** The following theorems give connections between the existence of points conjugate to a point A and the topological properties of the n-dimensional manifolds to which the n-dimensional element around A can be continued.

THEOREM 8. *Given an n-dimensional element E around a point A. Suppose that the coefficients a_{ij} in the Riemannian quadratic form for coördinates (y) along the neighborhood of an arbitrary geodesic g issuing from A are analytic functions of (y) for $(y_1, \cdots, y_{n-1})$ in the neighborhood of their values on g and for $0 < y_n < \infty$. Then if A has no conjugate point, the element E can be continued to a complete n-dimensional manifold homeomorphic to n-dimensional space.*

Since A has no conjugate point, according to Theorem 6 the determinant $|a_{ij}|$ is different from zero for $(y_1, \cdots, y_{n-1})$ near their values on g and for $0 < y_n < \infty$, and hence the quadratic form $a_{ij}dy_idy_j + dy_n^2$ will be positive definite throughout this region of the variables (y).

These coördinate systems along the various geodesics through A overlap, and in a region of overlapping in E one coördinate system can be obtained from the other by an admissible transformation of coördinates keeping invariant the

[18] See Blaschke, *Vorlesungen über Differentialgeometrie*, vol. I, 1930, p. 152.

Riemannian quadratic form, since all these coördinate systems were got directly by admissible transformations from one original coördinate system in the element E.

Now the geodesics through A in the element E can be put into a one-to-one correspondence with the straight lines through a point in n-space. The initial segment of each such straight line ray and its neighborhood can be provided with the Riemannian geometry belonging to the corresponding geodesic through A and its neighborhood in E. By the remarks in the first paragraph of this proof, this Riemannian geometry can be extended all along the neighborhood of the straight line. The relation between the coördinate systems of two such overlapping neighborhoods is determined by the corresponding overlapping in E.

The result is an n-dimensional analytic Riemannian manifold M containing the element E, and homeomorphic to n-space. The manifold M is complete; for every bounded set of points lies within the interior of a geodesic hypersphere of finite radius about A, and hence has a limit point on M.

THEOREM 9. *If a complete simply-connected n-dimensional manifold contains a point without conjugate point, then it is homeomorphic to n-dimensional space.*

Let the manifold be called M and the point without conjugate point be called A. Then by Theorem 6 the determinant $| a_{ij} |$ formed for coördinates (y) in the neighborhood of an arbitrary geodesic arc issuing from A does not vanish. Hence, by the corollary to Theorem 7, the hypotheses of Theorem 8 are satisfied, and the element of M around A can be continued to a complete n-dimensional manifold homeomorphic to n-dimensional space. From the uniqueness theorem (Theorem 3) we deduce that M is homeomorphic to n-dimensional space.

COROLLARY 1. *If a complete n-dimensional manifold M contains a point A without conjugate point, then the universal covering manifold of M is homeomorphic to n-dimensional space.*

For the element around A can be continued to the universal covering manifold of M, which will be complete. An application of Theorem 10 proves the corollary.

COROLLARY 2. *Through every point A of a complete n-dimensional manifold whose universal covering manifold is not homeomorphic to n-dimensional space passes a geodesic containing a point conjugate to A. In particular, through every point A of a simply-connected, closed, n-dimensional manifold passes a geodesic containing a point conjugate to A.*

THEOREM 10. *Through every point A of a complete open n-dimensional manifold M passes a geodesic which is the shortest line between A and any of its points, and hence contains no point conjugate to A. Through every point A of a closed n-dimensional manifold M whose universal covering manifold is open passes a geodesic containing no point conjugate to A.*

If M is open, it will contain a sequence of points P_i without a limit point. Let A be an arbitrary point on M. Since M is complete, the distance $\rho(A, P_i)$ from A to P_i becomes infinite as i becomes infinite. Furthermore, we can pass a geodesic c_i of shortest length between A and P_i. Let $\rho(A, P_i) = r_i$.

Interior to a neighborhood of A simply covered by the geodesics through A, consider a geodesic hypersphere H. Corresponding to the sequence of points P_i and geodesics c_i there will be a sequence of points P_i' on H at which the geodesics c_i intersect H. The sequence P_i' will have a limit point P' on H. Let c be the geodesic AP', Q any point on c, and L the length of AQ on c. A certain subsequence $\overline{P}_i'$ of P_i' will converge to P'. Let $\bar{c}_i$ be the geodesic joining A to $\overline{P}_i'$, and $\overline{P}_i$ be the subsequence of P_i on the geodesics $\bar{c}_i$. Then for sufficiently large i the lengths $A\overline{P}_i = \bar{r}_i$ on the geodesics $\bar{c}_i$ will be greater than L.

Measure off the length L on the geodesics $\bar{c}_i$ thus getting points $\bar{Q}_i$. Since $\bar{c}_i$ is a geodesic of shortest length between A and $\overline{P}_i$, and since $\bar{r}_i > L$, $\bar{c}_i$ is also a shortest line from A to $\bar{Q}_i$. Hence $\rho(A\bar{Q}_i) = L$.

The points $\bar{Q}_i$ converge to the point Q, because of the continuous dependence of the geodesics through A on the points of H through which they pass. Therefore $\rho(A, \bar{Q}_i) \rightarrow \rho(A, Q)$, and hence $\rho(A, Q) = L$. But L was the length of AQ on c, so that we have proved that c is the shortest line from A to Q.

But Q was any point on c. Thus c is the shortest line from A to any of its points, and contains no point conjugate to A.

The second part of the theorem is proved as follows. If M is a closed n-dimensional manifold whose universal covering manifold M' is open, then we can apply the first part of the theorem to M'. Hence through every point A' of M' passes a geodesic c' without a point conjugate to A'. Now M' can be mapped on M in a single-valued, im kleinen isometric manner. If A is the point of M corresponding to A', and c the geodesic corresponding to c', then c can contain no point conjugate to A. Since every point A of M corresponds to some point A' of M', the theorem is proved.

COROLLARY 1. *If a complete n-dimensional manifold M contains a point A such that every geodesic through A contains a point conjugate to A, then the universal covering manifold of M is closed.*

COROLLARY 2. *An n-dimensional element E about a point A such that on every geodesic through A there exists a point conjugate to A cannot be continued to a complete open n-dimensional manifold.*

HARVARD UNIVERSITY.

ON THE HOMOLOGY CHARACTERS OF SYMMETRIC PRODUCTS

By Moses Richardson

In the first part of this paper, we consider a complex K, subjected to the transformations of a group of finite order p. We define a new complex k, the so-called "domain of discontinuity" of the group, by identifying all points which are images of each other under the transformations of the group. We then determine the Betti numbers of k, both non-modular and mod π^u, π a prime and not a factor of p. In the second part, we use the results of Part I to obtain explicit formulae for the Betti numbers of 2-fold and 3-fold symmetric products of a complex in terms of its own Betti numbers, and we indicate a general procedure by which the Betti numbers of a q-fold symmetric product can be computed. In the third part, we continue the development of the theory of the 2-fold symmetric product. The methods of Part II yield the Betti numbers mod π^u when π is an odd prime but not when $\pi = 2$. The Betti numbers mod 2 have been found by P. A. Smith.[1] By an extension of his methods, we determine the Betti numbers mod 2^u, $u > 1$. With a knowledge of all the Betti numbers modulo powers of primes, the torsion coefficients can be determined.[2]

Part I

1. Consider a simplicial oriented n-complex K. Let G be a group, of finite order p, of (1, 1) continuous transformations of K into itself. The transformations $I = T_0, T_1, \cdots, T_{p-1}$ of G will be subject to the following restrictions:

(a) they carry m-simplexes of K into m-simplexes of K;

(b) no T_λ $(\lambda = 0, 1, \cdots, p-1)$ carries a vertex into an adjacent[3] vertex.

As a result of (b), a simplex can be invariant only if it is pointwise invariant.

If $E_m = V_0V_1 \cdots V_m$ is an oriented[4] m-simplex of K, we define $T_\lambda E_m$ to be the oriented simplex $T_\lambda V_0 T_\lambda V_1 \cdots T_\lambda V_m$ of K. Likewise, if C is the chain $\Sigma t_i E^i_m$, we define $T_\lambda C$ to be the chain $\Sigma t_i T_\lambda E^i_m$. Since

$$E_m \to \sum_{s=0}^{m} (-1)^s V_0 \cdots V_{s-1} V_{s+1} \cdots V_m$$

Received by the Editors of the Annals of Mathematics, May 25, 1934, accepted by them, and later transferred to this journal.

[1] P. A. Smith, *The topology of involutions*, Proc. Nat. Acad. Sci., vol. 19 (1933), pp. 612–618. The author is indebted to Prof. Smith for much stimulating guidance received during the preparation of this paper.

[2] A. W. Tucker, *Modular homology characters*, Proc. Nat. Acad. Sci., vol. 18 (1932), pp. 467–471.

[3] Two vertices are called adjacent if they are distinct and if both occur in the symbol of the same simplex.

[4] Our notation and terminology will be largely that of S. Lefschetz, *Topology*, Amer. Math. Soc. Colloquium Publications, No. 12, New York, 1930.

and

$$T_\lambda E_m \to \sum_{s=0}^{m} (-1)^s T_\lambda V_0 \cdots T_\lambda V_{s-1} T_\lambda V_{s+1} \cdots T_\lambda V_m ,$$

it is clear that $T_\lambda F(E_m) = F(T_\lambda E_m)$. Hence $T_\lambda F(C) = F(T_\lambda C)$.

2. To each of the points $V_i, T_1 V_i, \cdots, T_{p-1} V_i$ we shall associate the mark v_i and shall write $\Lambda T_\lambda V_i = v_i$ ($\lambda = 0, 1, \cdots, p-1$). Suppose V_i and V_j are such that $V_i \neq T_\lambda V_j$ ($\lambda = 0, 1, \cdots, p-1$); in particular, this condition is satisfied whenever V_i and V_j are adjacent, by (b), §1. Then the sets

$$(V_i, T_1 V_i, \cdots, T_{p-1} V_i)$$

and

$$(V_j, T_1 V_j, \cdots, T_{p-1} V_j)$$

have no common element; hence $v_i \neq v_j$. It follows that if $V_0, V_1, \cdots, V_m$ are the vertices of a non-oriented simplex $| E_m | = | V_0 V_1 \cdots V_m |$, then $v_0, v_1, \cdots, v_m$ are distinct and can therefore be taken as the vertices of an m-simplex. If E_m is the oriented simplex $V_0 V_1 \cdots V_m$, we define ΛE_m to be the oriented simplex $\Lambda V_0 \Lambda V_1 \cdots \Lambda V_m = v_0 v_1 \cdots v_m = e_m$, say. If $| V_{i_0} V_{i_1} \cdots V_{i_k} |$ is a k-face of $| E_m |$, then it is clear that $| v_{i_0} v_{i_1} \cdots v_{i_k} |$ is a k-face of $| e_m |$; hence the totality of oriented simplexes $e_m^i = \Lambda E_m^i$ ($m = 0, 1, \cdots, n$; $i = 1, 2, \cdots$) constitutes an oriented n-complex which we shall denote by k. Our purpose is to determine the Betti numbers of k.

If $C = \Sigma t_i E_m^i$, we define ΛC to be the chain $c = \Sigma t_i \Lambda E_m^i = \Sigma t_i e_m^i$. Since $e_m \to \sum_{s=0}^{m} (-1)^s v_0 \cdots v_{s-1} v_{s+1} \cdots v_m$, it is clear that $F(\Lambda E_m) = \Lambda F(E_m)$. Consequently, Λ *preserves bounding relations.* Thus, for example, if $C \sim 0$, then $\Lambda C \sim 0$.

Let e_m be an oriented m-simplex of k. We have

$$e_m = \Lambda T_\lambda E_m \ (\lambda = 0, 1, \cdots, p-1).$$

We now define $\Lambda' e_m$ to be the chain $\sum_{\lambda=0}^{p-1} T_\lambda E_m$ on K. And if $c = \Sigma t_i e_m^i$, then we define $\Lambda' c$ to be the chain $\Sigma t_i \Lambda' e_m^i$ on K. It is easily seen that if $e_m \to \Sigma t_i e_{m-1}^i$, then $\Lambda' e_m \to \Sigma t_i \Lambda' e_{m-1}^i$. Thus Λ' *also preserves bounding relations.* It is to be noted that Λ and Λ' are not the inverses of each other. We have instead the relations

$$\Lambda' \Lambda C = \sum_{\lambda=0}^{p-1} T_\lambda C, \tag{2.1}$$

$$\Lambda \Lambda' c = pc. \tag{2.2}$$

Note that everything said so far may be understood to be mod q without any change. For example, Λ and Λ' preserve bounding relations mod q.

3. Let $\Gamma^1, \Gamma^2, \cdots, \Gamma^s$ be the cycles of a minimal base with respect to homology mod π^u, π a prime, of m-cycles of K. All homologies in this section are understood to be mod π^u, and all equations are understood to be congruences mod π^u. Let $\Gamma^1, \Gamma^2, \cdots, \Gamma^r$, $r = R_m(K)$, be the elements of a minimal base for weak homology[5] and $\Gamma^{r+1}, \Gamma^{r+2}, \cdots, \Gamma^s$ the elements of a base for zero-divisors. Note that Γ^i, $i \leqq r$, is an element of order[6] π^u, while Γ^i, $i > r$, is an element of order π^a, $a < u$. We have

$$(3.1) \qquad T_\lambda \Gamma^i_m \sim \sum_{j=1}^{s} {}^m a^\lambda_{i\,j}\, \Gamma^j_m \qquad (i = 1, 2, \cdots, s; \lambda = 0, 1, \cdots, p-1),$$

where the determinant of each matrix $({}^m a^\lambda_{i\,j})$ is ± 1 and the matrix $({}^m a^0_{i\,j})$ is in fact the identity matrix (δ_{ij}) of order s. Consider the matrix $(x^m_{i\,j}) = \sum_{\lambda=0}^{p-1} ({}^m a^\lambda_{i\,j})$. We may suppose that the first ρ rows of $(x^m_{i\,j})$ are linearly independent mod π^u, where $\rho = \rho(x^m)$ is the maximum number of rows, or columns,[7] of $(x^m_{i\,j})$ linearly independent mod π^u. Note that $\rho \leqq r$ since the zero-divisors cannot contribute independent rows.

(3.2) *If* $(s_{ik}) = (t_{ij})(u_{jk})$, *then* $\rho(s) \leqq \rho(u)$. For, suppose that

$$s_{i1}, s_{i2}, \cdots, s_{i,\rho+1},$$

where $\rho = \rho(u)$, were independent columns. We can pick integers

$$c_1, c_2, \cdots, c_{\rho+1},$$

not all congruent to zero, such that $\sum_{k=1}^{\rho+1} c_k u_{jk} = 0$ for all values of j. Then

$$\sum_{k=1}^{\rho+1} c_k s_{ik} = \sum_{k=1}^{\rho+1} c_k \sum_j t_{ij} u_{jk} = \sum_j \left(\sum_{k=1}^{\rho+1} c_k u_{jk} \right) t_{ij} = 0$$

for all values of i.

We can now prove the

THEOREM 1. *If p is not divisible by π, then* $R_m(k, \pi^u) = \rho(x^m)$.

Proof. (A). There exist ρ m-cycles mod π^u of k, namely, the cycles $\gamma^1 = \Lambda\Gamma^1$, $\gamma^2 = \Lambda\Gamma^2, \cdots, \gamma^\rho = \Lambda\Gamma^\rho$, which are linearly independent with respect to homology mod π^u. For, suppose that $\sum_{i=1}^{\rho} t_i \gamma^i \sim 0$, where some $t_i \not\equiv 0$. Then $\sum_{i=1}^{\rho} t_i \Lambda' \gamma^i \sim 0$, or, by (2.1), $\sum_{i=1}^{\rho} t_i \sum_{\lambda=0}^{p-1} T_\lambda \Gamma^i \sim 0$. By (3.1) we have

$$\sum_{i=1}^{\rho} t_i \sum_{\lambda=0}^{p-1} {}^m a^\lambda_{ij} \Gamma^j \sim 0,$$

[5] Following P. Alexandroff *Einfachste Grundbegriffe der Topologie*, Berlin, 1932, we shall say C is "weakly homologous to zero" when $C = 0$.

[6] The order of a cycle C is the least positive integer t such that $tC \sim 0$.

[7] The maximum number of rows linearly independent mod π^u is equal to the maximum number of columns linearly independent mod π^u, since reduction of the matrix to canonical form preserves both numbers and is unique. Cf. J. W. Alexander, *Combinatorial analysis situs*, Trans. Amer. Math. Soc., vol. 28 (1926), pp. 301–329.

or $\sum_{i=1}^{\rho} t_i x_{ij}^m \Gamma^j \sim 0$ which implies that $\sum_{i=1}^{\rho} t_i x_{ij}^m = 0$ for all values of j. But this contradicts the hypothesis that the first ρ rows of (x_{ij}^m) are independent.

(B). Any $\rho + 1$ arbitrary m-cycles mod π^u, say $\delta^1, \delta^2, \cdots, \delta^{\rho+1}$, of k are linearly dependent with respect to homology mod π^u. For we can write $\Lambda'\delta^i = D^i = \sum_{\lambda=0}^{p-1} T_\lambda \Delta^i$, where Δ^i is a chain such that $\Lambda\Delta^i = \delta^i$. Since D^i is a cycle, we have $D^i \sim \sum_{j=1}^{s} y_{ij}\Gamma^j$ $(i = 1, 2, \cdots, \rho + 1)$. Now, by (3.1),

$$T_\lambda D^i \sim \sum y_{ij} T_\lambda \Gamma^j \sim \sum_{j,k} y_{ij}\, {}^m a_{jk}^\lambda \Gamma^k .$$

Thus,

$$\sum_{\lambda=0}^{p-1} T_\lambda D^i \sim \sum_{j,k} y_{ij} \sum_{\lambda=0}^{p-1} {}^m a_{jk}^\lambda \Gamma^k = \sum_{j,k} y_{ij} x_{jk}^m \Gamma^k = \sum_k z_{ik} \Gamma^k, \text{ say.}$$

Since $T_\lambda D^i = D^i$ we have $pD^i \sim \Sigma z_{ik}\Gamma^k$. Since $z_{ik} = \sum_j y_{ij} x_{jk}^m$ we can find integers $t_1, t_2, \cdots, t_{\rho+1}$, not all zero, such that $\sum_1^{\rho+1} t_i z_{ik} = 0$ for all values of k, by (3.2). Therefore $p\,\Sigma t_i D^i \sim \Sigma t_i z_{ik} \Gamma^k = 0$. Hence, operating with Λ, we have $p \Sigma t_i \Lambda D^i \sim 0$ or $p^2 \Sigma t_i \delta^i \sim 0$. Let t be the g.c.d. of $t_1, t_2, \cdots, t_{\rho+1}$. The cycles $\delta^1, \delta^2, \cdots, \delta^{\rho+1}$ are dependent provided that $p^2 t \neq 0$. In particular, since $t \neq 0$, this condition is satisfied if $p \not\equiv 0 \bmod \pi$. This completes the proof.

The following two corollaries are immediate.

(3.3) *If p is not divisible by π, and $R_m(K, \pi^u) = 0$, then $R_m(k, \pi^u) = 0$.*

(3.4) *If p is not divisible by π, and K is an n-sphere, then*

$$R_m(k, \pi^u) = 0 \qquad (m = 1, 2, \cdots, n - 1).$$

(3.5) In the non-modular case we need not consider the entire base

$$\Gamma^1, \Gamma^2, \cdots, \Gamma^s$$

for homology of m-cycles. It will in fact be more convenient in the applications to follow in Part II of this paper to use only the cycles $\Gamma^1, \Gamma^2, \cdots, \Gamma^r$, $r = R_m(K)$, which form a minimal base for weak homology, and to employ the symbol $\approx$ throughout. Thus, we write $T_\lambda \Gamma_m^i \approx \sum_{j=1}^{r} {}^m a_{ij}^\lambda \Gamma_m^j$, $(i = 1, 2, \cdots, r)$ and $(x_{ij}^m) = \sum_{\lambda=0}^{p-1} ({}^m a_{ij}^\lambda)$. Then, *in the non-modular case, $R_m(k) = \rho(x^m)$ where ρ is the ordinary rank of (x_{ij}^m).* The proof follows exactly the same lines as that of Theorem 1. The corollaries analogous to (3.3) and (3.4)[8] also hold in the non-modular case.

[8] The corollary analogous to (3.4) in the non-modular case was found by Threlfall and Seifert, *Topologische Untersuchung der Diskontinuitätsbereiche endlichen Bewegungsgruppen des dreidimensionalen sphärischen Raumes*, Math. Annalen, vol. 104 (1931), pp. 1–70, for the special case where $n = 3$ and G is a finite group of rotations.

(3.6) Let $R_m(K, \pi) = 0$ for all primes π. Let N be a prime such that the number $t_m(N)$ of m-dimensional coefficients of torsion of k divisible by N is not zero. By the well known relation

$$R_m(k, N) = R_m(k) + t_m(N) + t_{m-1}(N),$$

we have $R_m(k, N) \neq 0$. By (3.3), this implies $p \equiv 0 \bmod N$. This proves that *if $R_m(K, \pi) = 0$ for all primes π, then every m-dimensional torsion coefficient of k is of the form $\pi_1^{a_1}\pi_2^{a_2} \cdots \pi_q^{a_q}$, where $\pi_1, \pi_2, \cdots, \pi_q$ are the prime factors of p. In particular this conclusion is valid when K is an n-sphere for*

$$m = 1, 2, \cdots, n - 1.$$ [9]

(3.7) Let K_n be an n-circuit. If G contains a transformation which reverses orientation, then the totality of orientation-preserving transformations of G forms an invariant subgroup of index 2. Thus half of the transformations of G reverse orientation. Now if $R_n(K_n) = 1$, then $R_n(k_n) = 0$, for $(x_{ij}^n) = (x_{11}^n) = (1 - 1 + 1 - 1 \cdots) = (0)$. If, however, G has no orientation-reversing transformation, and if $R_n(K_n) = 1$, then $R_n(k_n) = 1$. If p is odd, then k is a circuit for G has no orientation-reversing transformation and $R_n(k_n, 2) = 1$. If p is odd and K_n is a non-orientable circuit, then so is k_n.

Part II

In this part we shall apply the results of Part I to find certain Betti numbers of symmetric products.

4. **Direct and symmetric products.** If A is a class of abstract elements, the direct q-fold product $A \times A \times \cdots \times A$ (q factors) is the set of ordered q-tuples $(x \times y \times \cdots \times z)$ where $x, y, \cdots, z$ are elements of A. The q-fold symmetric product of A is the set of non-ordered q-tuples $(x, y, \cdots, z)$ where $x, y, \cdots, z$ are elements of A. The symmetric q-fold product of A can evidently be obtained from the direct q-fold product by merely identifying the element $(x \times y \times \cdots \times z)$ of $A \times A \times \cdots \times A$ with all its images under the group of permutations of the $x, y, \cdots, z$.

5. **Subdivision of direct q-fold product complexes.** Let K_n be a complex of simplexes in a cartesian space S_ν. The direct q-fold product

$$K_n \times K_n \times \cdots \times K_n = K_{qn}$$

can be considered as being immersed in a cartesian space

$$S_\nu \times S_\nu \times \cdots \times S_\nu = S_{q\nu},$$

and will evidently be composed of flat convex cells. Let any point X of K_n have the coördinates $x_1, x_2, \cdots, x_\nu$, in S_ν. The coördinates in $S_{q\nu}$ of the point

[9] See note 8.

$$X \times Y \times \cdots \times Z \tag{5.1}$$

of K_{qn} are $(x_1, x_2, \cdots, x_\nu, y_1, y_2, \cdots, y_\nu, \cdots, z_1, z_2, \cdots, z_\nu)$, or in condensed notation

$$(x, y, \cdots, z). \tag{5.2}$$

Now we take K_{qn} as our basic complex, and a group isomorphic to the symmetric group on q letters as our group G: namely, the group which takes the point (5.1) with coördinates (5.2) into all points obtainable by permuting the letters in these symbols. All the points of K_{qn} of the form

$$X \times X \times \cdots \times X \tag{5.3}$$

are invariant under the $q!$ transformations of G, and the totality of them constitutes a point set homeomorphic to K_n. Likewise the set I of points (5.3) contained in a cell of K_{qn} of the type

$$E^i_\mu \times E^i_\mu \times \cdots \times E^i_\mu \tag{5.4}$$

is homeomorphic to E^i_μ; and only cells of type (5.4) contain points of the type (5.3). Thus I is a μ-cell which we shall call E^{0i}_μ, and the totality of these cells constitutes a complex homeomorphic to K_n which we shall call K^0_n.

The cells of K_{qn} are merely permuted by the transformations T_λ of G. We shall now obtain a simplicial subdivision $K^{(1)}_{qn}$ of K_{qn} which will satisfy the requirements (a) and (b) of §1 with respect to the T_λ's. The process of subdivision begins with the 1-cells and proceeds to the m-cells only after all the $(m - 1)$-cells have been subdivided. The procedure for m-cells is as follows.

Let $|\sigma^1|, |\sigma^2|, \cdots$, be the $(m-1)$-simplexes on the boundary of the m-cell E of K_{qn}. Now: (1) if E is of type (5.4) we introduce an arbitrary point P, situated on the μ-cell of K^0_n which is imbedded in E, as a new vertex; then we replace E by the set of all m-simplexes $|P\sigma^1|, |P\sigma^2|, \cdots$, together with all the faces of lower dimension of these new m-simplexes except those on $F(E)$; (2) if E is not of type (5.4), we introduce an arbitrary point P in E as a new vertex, and proceed as in (1); simultaneously we introduce the points

$$T_\lambda P \; (\lambda = 0, 1, \cdots, p - 1)$$

in $T_\lambda E$ as new vertices, and we subdivide the cells $T_\lambda E$ similarly. The subdivision thus obtained is called $K^{(1)}_{q\,n}$.

$K^{(1)}_{q\,n}$ satisfies requirements (a) and (b) of §1 with respect to the T's.

Proof. (a) Any simplex $|E_m| = |V_0V_1 \cdots V_m|$ is flat, and the T's are evidently linear in the coördinates of $S_{q\nu}$. Thus $T_\lambda E_m$ is also flat and is therefore completely determined by its vertices $T_\lambda V_0, T_\lambda V_1, \cdots, T_\lambda V_m$.

(b) The vertices V and $T_\lambda V$ of $K^{(1)}_{q\,n}$ cannot be adjacent. (1) Suppose V and $T_\lambda V$ are vertices of K_{qn}. If they had been joined by a 1-cell, they were separated by the insertion of a new vertex on that 1-cell; and if there had been no such 1-cell, they are certainly not adjacent. (2) If V is not a vertex of K_{qn}, then

neither is $T_\lambda V$. Now, when V and $T_\lambda V$ were introduced within the convex cells of K_{qn} containing them, they were joined by new 1-cells only to the vertices of these containing cells; thus there can be no 1-cell of $K^{(1)}_{q\,n}$ joining V and $T_\lambda V$.

Now we have a simplicial subdivision $K^{(1)}_{q\,n}$ of K_{qn} which satisfies all the requirements of Part I. Each convex cell of K_{qn} has been replaced by a set of simplexes which we shall orient concordantly with the original convex cell. In an m-chain, we shall have each m-simplex of $K^{(1)}_{q\,n}$ affected by the same coefficient as affected the original convex cell of K_{qn} previous to the subdivision.

Evidently, $\Lambda K^{(1)}_{q\,n} = k_{qn}$ *is the q-fold symmetric product of* K_n.

(5.5) The closure of each cell of $K^{(1)}_{q\,n}$ is a subcomplex of $K^{(1)}_{q\,n}$. Let some of these subcomplexes be subdivided by section[10] into convex complexes. It is clear that we can further subdivide $K^{(1)}_{q\,n}$ into a complex of simplexes, say $K^{(2)}_{q\,n}$, which will again satisfy requirements (a) and (b) of §1.

6. **Two-fold symmetric products.** Let K_n be a simplicial complex, $K_{2n} = K_n \times K_n$ the direct product complex, $K^{(1)}_{2\,n}$ the subdivision of K_{2n}, and S_ν the space in which K_n is immersed (§5). The group G has the elements I, T where T is the transformation which interchanges the points $X \times Y$ and $Y \times X$ of K_{2n}. To obtain the m^{th} Betti number $R_m(k_{2n})$ of the 2-fold symmetric product complex $k_{2n} = \Lambda K^{(1)}_{2\,n}$, we have only to write down explicitly the matrix $(x^m_{i\,j}) = ({}^m a_{ij}) + (\delta_{ij})$ mentioned in (3.5) and to compute its rank.

Let $E_p = A_0 A_1 \cdots A_p$ and $E_q = B_0 B_1 \cdots B_q$ be simplexes of K_n. Let A_i have the coördinates $(x^i_1, x^i_2, \cdots, x^i_p)$, denoted briefly by x^i, in the cartesian subspace S_p of S_ν determined by $A_0, A_1, \cdots, A_p$. Likewise, let B_j have the coördinates $(y^j_1, y^j_2, \cdots, y^j_q)$, denoted by y^j, in the subspace S_q of S_ν determined by the vertices of E_q.

Lemma. $T(E_p \times E_q) = (-1)^{pq} E_q \times E_p$.

Proof. The orientation of the cell $E_p \times E_q$ is uniquely determined by the sign of the determinant of the homogeneous coördinates ($x_0 = 1$) of the vertices of the simplex[11]

$$A_0 \times B_0\, A_1 \times B_0 \cdots A_p \times B_0\, A_0 \times B_1 \cdots A_0 \times B_q. \tag{6.1}$$

The orientation of $E_q \times E_p$ is likewise determined by the simplex

$$B_0 \times A_0\, B_1 \times A_0 \cdots B_q \times A_0\, B_0 \times A_1 \cdots B_0 \times A_p. \tag{6.2}$$

Now the transform $T(E_p \times E_q)$ of $E_p \times E_q$ has its orientation determined by the transform of the simplex (6.1), namely

$$B_0 \times A_0\, B_0 \times A_1 \cdots B_0 \times A_p\, B_1 \times A_0 \cdots B_q \times A_0. \tag{6.3}$$

The determinants of the coördinates of (6.2) and (6.3) in $S_p \times S_q$ are, in condensed notation,

[10] Lefschetz, loc. cit., p. 67.

[11] Lefschetz, loc. cit., p. 224.

$$\begin{vmatrix} 1 & y^0 & x^0 \\ 1 & y^1 & x^0 \\ \cdot\cdot & \cdot\cdot & \cdot\cdot \\ 1 & y^q & x^0 \\ 1 & y^0 & x^1 \\ \cdot\cdot & \cdot\cdot & \cdot\cdot \\ 1 & y^0 & x^p \end{vmatrix} \quad \text{and} \quad \begin{vmatrix} 1 & y^0 & x^0 \\ 1 & y^0 & x^1 \\ \cdot\cdot & \cdot\cdot & \cdot\cdot \\ 1 & y^0 & x^p \\ 1 & y^1 & x^0 \\ \cdot\cdot & \cdot\cdot & \cdot\cdot \\ 1 & y^q & x^0 \end{vmatrix}.$$

We can evidently transform one of these determinants into the other by pq interchanges of rows, which proves the lemma. An immediate corollary is the corresponding relation for chains, i.e., $T(C_p \times C_q) = (-1)^{pq} C_q \times C_p$. By §5, this formula holds for $K_{2n}^{(1)}$.

Consider the set of cycles a_s^i $(i = 1, 2, \cdots, R_s(K_n))$ constituting a minimal base for weak homology of s-cycles of K_n. The set of cycles $a_s^i \times a_{m-s}^j$ form a minimal base[12] for weak homology of m-cycles on K_{2n}. If $m = 2s$, there are some m-cycles of the form $a_s^i \times a_s^i = \Delta_{2s}^i$, which are transformed by T into themselves, except perhaps for orientation, since $T(a_s^i \times a_s^i) = (-1)^s a_s^i \times a_s^i$. These are evidently $R_s(K_n)$ in number. All other cycles of the base are transformed by T into different cycles of the base. We can therefore rename the cycles of the base as follows:

$$\Gamma^1, \Gamma^2, \cdots, \Gamma^g, \bar{\Gamma}^1, \bar{\Gamma}^2, \cdots, \bar{\Gamma}^g, \Delta^1, \Delta^2, \cdots, \Delta^h \tag{6.4}$$
$$(h = R_s(K_n);\ h + 2g = R_m(K_{2n})),$$

where $\bar{\Gamma}^i = T\Gamma^i$, $T\Delta^i = \bar{\Delta}^i = (-1)^s \Delta^i$, and $h = 0$ unless $m = 2s$. We are now able to write down the matrix of the weak homologies $T\Gamma_m^i \approx \Sigma\, {}^m a_{ij} \Gamma_m^j$, where we shall arrange the cycles in our matrix table in the order (6.4) horizontally and vertically beginning with the upper left hand corner. If $m \neq 2s$, we have

$$({}^m a_{ij}) = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix},$$

and if $m = 2s$,

$$({}^m a_{ij}) = \begin{pmatrix} 0 & I & 0 \\ I & 0 & 0 \\ 0 & 0 & (-1)^s I' \end{pmatrix},$$

where I and I' represent identity matrices. Bearing in mind the orders of these submatrices, we compute easily the rank of the matrix $(x_{i\,j}^m) = ({}^m a_{ij}) + (\delta_{ij})$, which, by (3.5) is equal to $R_m(k_{2n})$. Thus,

[12] Lefschetz, loc. cit., p. 228.

THEOREM 2. *The non-modular m-dimensional Betti number of the 2-fold symmetric product k_{2n} of K_n is given by*

$$(6.5)\qquad R_m(k_{2n}) = \begin{cases} \frac{1}{2} R_m(K_{2n}) & (m \neq 2s) \\ \frac{1}{2}[R_m(K_{2n}) - R_s(K_n)] & (m = 2s,\ s \text{ odd}) \\ \frac{1}{2}[R_m(K_{2n}) - R_s(K_n)] + R_s(K_n) & (m = 2s,\ s \text{ even}) \end{cases}$$

where $K_{2n} = K_n \times K_n$.

Since $p = 2$, it is easily seen that *formulae* (6.5) *give the Betti numbers mod π^u provided π is an odd prime,* where all Betti numbers in the formulae are understood to be mod π^u.

Examples. (1) Let K_n be a 1-sphere H_1. For H_1 we have $R_0 = R_1 = 1$. For the torus $H_1 \times H_1$ we have $R_0 = R_2 = 1$, $R_1 = 2$. For the Möbius strip $\Lambda(H_1 \times H_1)$ we have $R_0 = \frac{1}{2}[1 - 1] + 1 = 1$, $R_1 = \frac{1}{2}[2] = 1$, $R_2 = \frac{1}{2}[1 - 1] = 0$. (2) Let K_n be a 2-sphere H_2. For H_2 we have $R_0 = R_2 = 1$, $R_1 = 0$. For $H_2 \times H_2$ we have $R_0 = R_4 = 1$, $R_1 = R_3 = 0$, $R_2 = 2$. For the complex projective plane $\Lambda(H_2 \times H_2)$ we have $R_0 = R_2 = R_4 = 1$, $R_1 = R_3 = 0$.

7. Two-fold symmetric product of a circuit. First, note that after the subdivision of §5, we can name the m-simplexes of the complex $K_{2n}^{(1)}$ as follows: $D_m^1, D_m^2, \cdots, D_m^\mu, \bar{D}_m^1, \bar{D}_m^2, \cdots, \bar{D}_m^\mu, D_m^{0\,1}, D_m^{0\,2}, \cdots, D_m^{0\,\nu}$, where $\bar{D}_m^i = TD_m^i$ and $D_m^{0\,i}$ denotes a simplex of the subcomplex K_n^0. The bounding relations of these simplexes can be written as follows:

$$(7.1)\qquad \begin{aligned} D_m^i &\to \Sigma\, \epsilon_{ij} D_{m-1}^j + \Sigma\, \eta_{ij} \bar{D}_{m-1}^j + \Sigma\, \zeta_{ij} D_{m-1}^{0\,j}, \\ \bar{D}_m^i &\to \Sigma\, \eta_{ij} D_{m-1}^j + \Sigma\, \epsilon_{ij} \bar{D}_{m-1}^j + \Sigma\, \zeta_{ij} D_{m-1}^{0\,j}, \\ D_m^{0\,i} &\to \Sigma\, \theta_{ij} D_{m-1}^{0\,j}, \end{aligned}$$

where the coefficients can take only the values $+1$, -1, or 0 and where not both ϵ_{ij} and η_{ij} are $\neq 0$.

On k_{2n} we have the bounding relations

$$(7.2)\qquad \begin{aligned} d_m^i &\to \Sigma\, (\epsilon_{ij} + \eta_{ij})\, d_{m-1}^j + \Sigma\, \zeta_{ij}\, d_{m-1}^{0\,j}, \\ d_m^{0\,i} &\to \Sigma\, \theta_{ij}\, d_{m-1}^{0\,j}, \end{aligned}$$

where $d^i = \Lambda D^i$ and $d^{0i} = \Lambda D^{0i}$, the d^{0i} being cells of $k_n^0 = \Lambda K_n^0$.

Now let K_n be an absolute n-circuit. The direct 2-fold product K_{2n} is an absolute $2n$-circuit which is orientable if and only if K_n is orientable.[13]

THEOREM 3. *If K_n is an absolute circuit, then, for $n = 1$, k_{2n} is a relative $2n$-circuit modulo its subcomplex k_n^0, and, for $n > 1$, k_{2n} is an absolute $2n$-circuit.*

Proof. On the subdivided complex $K_{2n}^{(1)}$, we have

$$\sum_i F(D_{2n}^i) + \sum_i F(\bar{D}_{2n}^i) \equiv 0 \bmod 2.$$

[13] Lefschetz, loc. cit., p. 231.

By (7.1) this becomes

$$\sum_{i\,j} [(\epsilon_{ij} + \eta_{ij})\, D^{j}_{2\,n-1} + (\epsilon_{ij} + \eta_{ij})\, \bar{D}^{j}_{2\,n-1} + 2\zeta_{ij}\, D^{0\,j}_{2\,n-1}] \equiv 0 \bmod 2\,.$$

Hence,

$$\sum_{i} F(d^{i}_{2\,n}) = \sum_{i\,j} [(\epsilon_{ij} + \eta_{ij})\, d^{j}_{2\,n-1} + \zeta_{ij}\, d^{0\,j}_{2\,n-1}] \equiv \sum_{i\,j} \zeta_{ij}\, d^{0\,j}_{2\,n-1} \bmod 2\,,$$

where evidently the cells $d^{0\,j}_{2\,n-1}$ exist only when $2n - 1 \leqq n$; i.e., when $n = 1$.[14] Thus k_{2n} is an absolute $2n$-cycle mod 2 for $n > 1$, and a relative $2n$-cycle mod 2 modulo k^{0}_{n} for $n = 1$.

We have yet to show that no subset of the $2n$-cells of k_{2n} forms a $2n$-cycle mod 2. Suppose there were a subset $d^{1}_{2\,n}, \cdots, d^{\sigma}_{2\,n}$ $(\sigma < \mu + \nu)$ such that $\sum_{i=1}^{\sigma} F(d^{i}_{2\,n}) \equiv 0$ mod 2, i.e.,

$$\sum_{i\,j} [(\epsilon_{ij} + \eta_{ij})\, d^{j}_{2\,n-1} + \zeta_{ij}\, d^{0\,j}_{2\,n-1}] \equiv 0 \bmod 2\,,$$

then it would follow that

$$\sum_{i=1}^{\sigma} F(\Lambda'\, d^{i}_{2\,n}) = F(D^{1}_{2\,n} + \cdots + D^{\sigma}_{2\,n} + \bar{D}^{1}_{2\,n} + \cdots + \bar{D}^{\sigma}_{2\,n}) \equiv 0 \bmod 2\,,$$

which would contradict the hypothesis that K_{2n} is a circuit. This proves the theorem.

THEOREM 4. *If the absolute circuit K_n is non-orientable, so is k_{2n}; if K_n is orientable, then k_{2n} is orientable or not according as n is even or odd.*

Proof. If K_n is non-orientable, $R_n(K_n) = 0$, so that, by (6.5), $R_{2n}(k_{2n}) = 0$. If K_n is orientable, $R_n(K_n) = 1$, so that, by (6.5), $R_{2n}(k_{2n}) = 1$ or 0 according as n is even or odd.[15]

Example. The Möbius strip is an example of Theorems 3 and 4 for $n = 1$. The "edge" of the strip is k^{0}_{1}. The strip is a relative 2-circuit modulo its edge, and is non-orientable.

8. **Three-fold symmetric products.** Let K_n be a simplicial n-complex, $K_{3n} = K_n \times K_n \times K_n$ the direct 3-fold product, $K^{(1)}_{3\,n}$ the subdivision of K_{3n} which must be made in order to satisfy the requirements of Part I, and $k_{3n} = \Lambda K^{(1)}_{3\,n}$ the 3-fold symmetric product. Here G is of course isomorphic with the symmetric group on 3 letters. We have

$$\begin{aligned}
T_0 E_p \times E_q \times E_r &= E_p \times E_q \times E_r\,,\\
T_1 E_p \times E_q \times E_r &= (-1)^{pq+pr}\, E_q \times E_r \times E_p\,,\\
T_2 E_p \times E_q \times E_r &= (-1)^{pr+qr}\, E_r \times E_p \times E_q\,,\\
T_3 E_p \times E_q \times E_r &= (-1)^{pq}\, E_q \times E_p \times E_r\,,\\
T_4 E_p \times E_q \times E_r &= (-1)^{qr}\, E_p \times E_r \times E_q\,,\\
T_5 E_p \times E_q \times E_r &= (-1)^{pq+pr+qr}\, E_r \times E_q \times E_p\,.
\end{aligned} \tag{8.1}$$

[14] The case $n = 0$ is obviously devoid of interest.
[15] Cf. (3 7).

The exponents of -1, which we shall call the *orientation exponents* of the transformations (8.1), are obtained by methods analogous to those of the Lemma of §6.

For example, let us obtain the orientation exponent of T_1, say. Let $E_p = A_0A_1 \cdots A_p$, $E_q = B_0B_1 \cdots B_q$, $E_r = C_0C_1 \cdots C_r$, and let the coördinates of A_i, B_j, C_k be, in condensed notation, x^i, y^j, z^k.

The simplex determining the orientation of $E_q \times E_r$ is

$$B_0 \times C_0B_1 \times C_0 \cdots B_q \times C_0B_0 \times C_1 \cdots B_0 \times C_r;$$

thus the simplex determining the orientation of $E_q \times E_r \times E_p$ is

$$\begin{aligned} &B_0 \times C_0 \times A_0B_1 \times C_0 \times A_0 \cdots B_q \times C_0 \times A_0B_0 \times C_1 \times A_0 \cdots \\ &\qquad B_0 \times C_r \times A_0B_0 \times C_0 \times A_1 \cdots B_0 \times C_0 \times A_p. \end{aligned} \tag{8.2}$$

The simplex determining the orientation of $T_1E_p \times E_q \times E_r$ is the T_1-transform of the simplex which determines the orientation of $E_p \times E_q \times E_r$. Writing out the latter simplex, and applying T_1 to it, we find the desired simplex to be

$$\begin{aligned} &B_0 \times C_0 \times A_0B_0 \times C_0 \times A_1 \cdots B_0 \times C_0 \times A_pB_1 \times C_0 \times A_0 \cdots \\ &\qquad B_q \times C_0 \times A_0B_0 \times C_1 \times A_0 \cdots B_0 \times C_r \times A_0. \end{aligned} \tag{8.3}$$

The determinants of the coördinates of (8.2) and (8.3) are

$$\begin{vmatrix} 1 & y^0 & z^0 & x^0 \\ 1 & y^1 & z^0 & x^0 \\ \cdot & \cdot & \cdot & \cdot \\ 1 & y^q & z^0 & x^0 \\ 1 & y^0 & z^1 & x^0 \\ \cdot & \cdot & \cdot & \cdot \\ 1 & y^0 & z^r & x^0 \\ 1 & y^0 & z^0 & x^1 \\ \cdot & \cdot & \cdot & \cdot \\ 1 & y^0 & z^0 & x^p \end{vmatrix}, \quad \text{and} \quad \begin{vmatrix} 1 & y^0 & z^0 & x^0 \\ 1 & y^0 & z^0 & x^1 \\ \cdot & \cdot & \cdot & \cdot \\ 1 & y^0 & z^0 & x^p \\ 1 & y^1 & z^0 & x^0 \\ \cdot & \cdot & \cdot & \cdot \\ 1 & y^q & z^0 & x^0 \\ 1 & y^0 & z^1 & x^0 \\ \cdot & \cdot & \cdot & \cdot \\ 1 & y^0 & z^r & x^0 \end{vmatrix}.$$

Since these determinants can be transformed into each other by $pq + pr$ interchanges of rows, we have the desired orientation exponent.

The relations (8.1) hold also for chains of $K^{(1)}_{3n}$. Now we consider the set of m-cycles $a^i_p \times a^j_q \times a^k_r (p + q + r = m)$ which constitutes a minimal base for weak homology of m-cycles on $K^{(1)}_{3n}$. Since the formulae (8.1) tell us the precise effect of any T_λ on any cycle of this base, we are able to write out the matrix (x^m_{ij}), and to compute its rank, as we did for the 2-fold symmetric product.

We omit the details of this process. The result is as follows:

THEOREM 5. *The m-dimensional non-modular Betti number of the 3-fold symmetric product k_{3n} of a simplicial complex K_n is given by*

$$R_m(k_{3n}) = \begin{cases} \frac{1}{3}R_m(K_{3n}) - R_1(K_n) - R_3(K_n) - \cdots - R_l(K_n) & (m \neq 3s), \\ \frac{1}{3}[R_m(K_{3n}) - R_s(K_n)] + R_s(K_n) - R_1(K_n) - R_3(K_n) - \cdots - R_l(K_n) \\ & (m = 3s), \end{cases}$$

where l is the highest odd integer $\leqq m/2$.

Since $p = 6$, *the same formulae yield the Betti numbers mod π^u provided π is a prime* > 3, where now all Betti numbers in these formulae are understood to be Betti numbers mod π^u.

Example. Let K_n be a 1-sphere H_1. On $K_{3n} = H_1 \times H_1 \times H_1$ we have $R_0 = R_3 = 1, R_1 = R_2 = 3$. On $k_{3n} = \Lambda(H_1 \times H_1 \times H_1)$, we have

$$R_0 = R_1 = 1, R_2 = R_3 = 0.$$

9. q-fold symmetric products. Consider a cell of the form

$$E_a \times E_b \times \cdots \times E_l \qquad (q \text{ factors}) \tag{9.1}$$

Let us call any transformation which interchanges two adjacent cell-symbols in (9.1) an "inversion." For example, the transformation that takes (9.1) into $E_b \times E_a \times \cdots \times E_l$ is an inversion. From a consideration of the structure of the determinants involved in the computation of the orientation exponents (§8), it is easily seen that any inversion applied to (9.1) will introduce a factor of -1 with an exponent equal to the product of the dimensions of the cell-symbols interchanged. Since any transformation of the symmetric group on q letters can be produced by successive inversions, its orientation exponent can be determined. It seems to be impossible to write a general expression for these exponents in terms of q; the difficulty of generalization seems to be at least partly group-theoretic. However, for a given integer $q > 1$, we can write out the transformations of the symmetric group on q letters, and compute the orientation exponents of each. These exponents give us exact knowledge of the effect of each T_λ upon any cycle. Now, given a definite complex K_n, we can write down the cycles of a base for $\approx$ of m-cycles on K_{qn}. Then we can write out the $q!$ matrices $({}^m a^\lambda_{ij})$ and compute the rank of $(x^m_{ij}) = \sum_{\lambda=0}^{q!-1} ({}^m a^\lambda_{ij})$. Thus, we can find the Betti numbers of the q-fold symmetric product k_{qn} for a given K_n and a given q.

Example. Let K_n be a 1-sphere and $q = 4$. Then on K_{4n} we have

$$R_0 = R_4 = 1, R_1 = R_3 = 4, R_2 = 6.$$

It is easy to pick out a basal set of cycles for each dimension. Computing the orientation exponents of the 24 transformations of G, writing out the 24 matrices $({}^m a_{ij}^{\lambda})$ for each dimension, and so on, we find that on k_{4n} we have

$$R_0 = R_1 = 1, R_2 = R_3 = R_4 = 0.$$

Part III

We shall now determine the mod 2^u Betti numbers $R_m(k_{2n}, 2^u)$ of the 2-fold symmetric product k_{2n} for $u > 1$. All homologies and equations in the following sections will be understood to be homologies and congruences mod 2^u, $u > 1$. Unless otherwise indicated, all cells, chains, etc., will be of dimension m.

10. Let $W^1, W^2, \cdots, W^r$ be the cycles of a base for weak homology of m-cycles and $Z^1, Z^2, \cdots, Z^s$, the cycles of a base for m-dimensional zero-divisors of a given complex K.

If $C^1, C^2, \cdots, D^1, D^2, \cdots$, etc., are chains of the same dimension, the symbol $[C, D, \cdots]$ shall denote a linear combination of these chains with integer coefficients not all zero. If the coefficients of such a linear combination are either 0 or 1, not all zero, we write $\{C, D, \cdots\}$. If the coefficients of a linear combination are allowed to be all zero, we write $[C, D, \cdots]'$ or $\{C, D, \cdots\}'$. If the same symbol appears more than once in an argument, it does not necessarily represent the same linear combination at each occurrence.

(10.1) *There can exist no relation of the form* $2X \sim \{W\} + [Z]'$.

Proof. Certainly $X \sim \Sigma A_i W^i + [Z]'$. Let t be the l.c.m. of the orders of the Z's. Since our modulus 2^u is a power of a prime, it is easily shown that t is not congruent to zero. Now

$$(10.2) \qquad 2tX \sim 2t(A_1 W^1 + A_2 W^2 + \cdots + A_r W^r).$$

Suppose there were a relation of the form $2X \sim \Sigma\, a_i W^i + [Z]'$ where the a's are 0 or 1, not all zero. Then

$$(10.3) \qquad 2tX \sim t(a_1 W^1 + a_2 W^2 + \cdots + a_r W^r).$$

Hence $2t \not\equiv 0$. From (10.2) and (10.3) we obtain

$$0 \sim t(2A_1 - a_1)W^1 + \cdots + t(2A_r - a_r)W^r.$$

Therefore $t(2A_i - a_i) = 0$ for all values of i. But some $a_i = 1$. Therefore some $2A_i - a_i$ is odd, and t must be congruent to zero. This is a contradiction. If there are no zero-divisors on the given complex, the proof is essentially the same.

11. From this point on all chains, etc., shall be of K_{2n} or the subdivision $K_{2n}^{(1)}$ (Part II) unless otherwise stated.

If a chain has[16] no cells of K_n^0, we shall attach an asterisk to its symbol, as X^*. If a chain has only cells of K_n^0, we attach a zero to its symbol as X^0.

[16] A chain is said to "have" a cell if the cell occurs in the symbol for the chain with a non-zero coefficient.

If $TX = \bar{X} = \epsilon X$, $|\epsilon| = 1$, the chain X will be called *invariant*. Sometimes an invariant chain will be called *positively* or *negatively invariant* according as $\epsilon = +1$ or -1. If $\bar{X} = X$, then X is of the form $X^* + \bar{X}^* + X^0$; and if $\bar{X} = -X$, then X is of the form $X^* - \bar{X}^*$. If $H \to C$, where H and C are both positively or both negatively invariant, we write $C \simeq 0$. These special homologies obey the same rules as ordinary homologies.

(11.1) *There can exist no relation of the form*

$$L + \bar{L} + \{W^0\} + [Z^0]' \simeq 0,$$

where the cycles W^{0i}, Z^{0i} form bases, as in §10, for K_n^0.

Proof. Suppose there were a relation

$$(11.2) \qquad H + \bar{H} + H^0 \to L + \bar{L} + \{W^0\} + [Z^0]' \simeq 0.$$

In any case we can write $L + \bar{L} = L^* + \bar{L}^* + 2L^0$, so that (11.2) becomes

$$(11.3) \qquad H + \bar{H} + H^0 \to L^* + \bar{L}^* + 2L^0 + \{W^0\} + [Z^0]'.$$

Let $H + \bar{H} \to L^* + \bar{L}^* + X^0$, say. If a cell of K_n^0 occurs in $F(H)$, it occurs in $F(\bar{H})$ with the same coefficient, so that $X^0 = 2Y^0$, say. Therefore

$$(11.4) \qquad H + \bar{H} \to L^* + \bar{L}^* + 2Y^0.$$

Let $L^0 - Y^0 = U^0$. Then from (11.3) and (11.4) we have

$$H^0 \to 2U^0 + \{W^0\} + [Z^0]' \simeq 0,$$

contradicting (10.1).

(11.5) *If $H + \epsilon\bar{H} \to C + \epsilon\bar{C}$, $|\epsilon| = 1$, then $F(C) \simeq 0$.*

Proof. Let $L = C - F(H)$. Then $L \to F(C)$. Now

$$F(C) + \epsilon F(\bar{C}) = F(F(H + \epsilon\bar{H})) = 0.$$

Thus $F(C)$ is invariant. Also

$$L + \epsilon\bar{L} = C + \epsilon\bar{C} - (F(H) + \epsilon F(\bar{H})) = 0.$$

Hence L is invariant. Therefore $F(C) \simeq 0$.

(11.6) *If $C + \bar{C} \simeq 0$, then the cycle ΛC either is $\sim$ on k_{2n} to a zero-divisor of k_n^0, or is ~ 0 on k_{2n}.*

Proof. By hypothesis, $H^* + \bar{H}^* + H^0 \to C + \bar{C}$. Let $H^* \to C - D$ and let $H^0 \to J^0$. Then $H^* + \bar{H}^* + H^0 \to C + \bar{C} - (D + \bar{D}) + J^0 = C + \bar{C}$. Hence $J^0 = D + \bar{D}$ and $2\Lambda D = \Lambda J^0$. But $\Lambda H^0 \to \Lambda J^0 \sim 0$ on k_n^0; therefore $2\Lambda D \sim 0$ on k_n^0. Thus ΛD either is a zero-divisor of k_n^0 or is ~ 0 on k_n^0. But $\Lambda H^* \to \Lambda C - \Lambda D$. Hence $\Lambda C \sim \Lambda D$ on k_{2n}, which completes the proof.

12. Let $a_s^i = \Sigma\, t_\alpha^i E_s^\alpha$ $(i = 1, 2, \cdots, R_s(K_n))$ be the cycles of a base for weak homology of s-cycles of K_n. Then

$$\Delta_{2s}^i = a_s^i \times a_s^i = \sum_{\alpha,\beta} t_\alpha^i t_\beta^i E_s^\alpha \times E_s^\beta = \sum_{\alpha \neq \beta} t_\alpha^i t_\beta^i E_s^\alpha \times E_s^\beta + \sum_\alpha (t_\alpha^i)^2 E_s^\alpha \times E_s^\alpha.$$

(12.1) *Associated with each* Δ^i_{2s} $(i = 1, 2, \cdots, R_s(K_n))$ *there exists on* $K^{(1)}_{2n}$ *or a suitable subdivision of* $K^{(1)}_{2n}$ *a sequence* $({}^i\Delta^{2s}_{2s}, {}^i\Delta^{2s}_{2s-1}, \cdots, {}^i\Delta^{2s}_{s})$ *of cycles and a sequence* $({}^iX^*_{2s}, {}^iX^*_{2s-1}, \cdots, {}^iX^*_{s+1})$ *of chains such that*

$${}^i\Delta^{2s}_j = {}^iX^*_j + \epsilon{}^i\bar{X}^*_j, \ |\epsilon| = 1; F({}^iX^*_j) = {}^i\Delta^{2s}_{j-1} \qquad (j = 2s, 2s-1, \cdots, s+1)$$

and ${}^i\Delta^{2s}_s = W^i_s + \overline{W}^i_s + W^{0\,i}_s$, *where the cycles* $W^{0\,i}_s$ *form a base for weak homology of s-cycles of* K^0_n.

Proof. Let us drop the superscript i temporarily and consider a definite Δ_{2s}. Let $E^\alpha_s \times E^\alpha_s$ be an invariant cell of K_{2n} occurring in the symbol for Δ_{2s} with a non-zero coefficient. Let S^α_{2s} be the flat subspace of $S_{2\nu}$ (§6) which contains $E^\alpha_s \times E^\alpha_s$ and let $(x^\alpha_1, x^\alpha_2, \cdots, x^\alpha_s, y^\alpha_1, y^\alpha_2, \cdots, y^\alpha_s)$ be a cartesian coördinate system in S^α_{2s} so chosen that T interchanges the points (x, y) and (y, x) of S^α_{2s}. Let $|\psi^\alpha_{2s}|$ be the set of non-oriented $2s$-simplexes by which the subdivision $K^{(1)}_{2n}$ replaces the non-oriented convex cell $|E^\alpha_s \times E^\alpha_s|$. The space S^α_{2s-1}, defined by the equation $x^\alpha_1 = y^\alpha_1$, contains all the invariant points in the closure Cl $|\psi^\alpha_{2s}|$ of $|\psi^\alpha_{2s}|$. It is clear that S^α_{2s-1} subdivides $|\psi^\alpha_{2s}|$ by section into two sets of non-oriented convex cells, $|\phi^\alpha_{2s}|$ and $|\bar{\phi}^\alpha_{2s}|$, which are interchanged by T and are such that the closure of each set is the closure of a $2s$-cell. Do this for every α. By (5.5) we can further subdivide $K^{(1)}_{2n}$ into a complex $K^{(2)}_{2n}$ of simplexes such that conditions (a) and (b) of §1 are satisfied with respect to T. Let ψ^α_{2s} and ϕ^α_{2s} be the chains obtained from $|\psi^\alpha_{2s}|$ and $|\phi^\alpha_{2s}|$ by taking the chain sum of their simplexes oriented concordantly with $E^\alpha_s \times E^\alpha_s$. For the sake of definiteness, let s be odd. Then ψ^α_{2s} is negatively invariant,[17] and $\psi^\alpha_{2s} = \phi^\alpha_{2s} - \bar{\phi}^\alpha_{2s}$. Now let

$$X^*_{2s} = \sum_{\alpha<\beta} t_\alpha t_\beta E^\alpha_s \times E^\beta_s + \sum t^2_\alpha \phi^\alpha_{2s}.$$

Then $\Delta_{2s} = X^*_{2s} - \bar{X}^*_{2s} = \Delta^{2s}_{2s}$, say. Since Cl $|\phi^\alpha_{2s}|$ is the closure of a $2s$-cell, $F(\phi^\alpha_{2s})$ is a simple circuit. Thus, in particular, the $2s$-simplexes of $F(\phi^\alpha_{2s})$ in S^α_{2s-1} occur with the coefficient one; hence, they occur in the symbol for $F(X^*_{2s})$ with the coefficient t^2_α. Let $\psi^\alpha_{2s-1} = F(\phi^\alpha_{2s})$ and let $|\psi^\alpha_{2s-1}|$ be the set of the non-oriented simplexes of ψ^α_{2s-1}. The space S^α_{2s-2}, defined by the equations $x^\alpha_1 = y^\alpha_1$, $x^\alpha_2 = y^\alpha_2$, contains all the invariant points in Cl $|\psi^\alpha_{2s}|$ and subdivides $|\psi^\alpha_{2s-1}|$ by section into two sets $|\phi^\alpha_{2s-1}|$ and $|\bar{\phi}^\alpha_{2s-1}|$ of non-oriented convex cells, which are interchanged by T, and are such that Cl $|\phi^\alpha_{2s-1}|$ is the closure of a $(2s-1)$-cell. Introduce this subdivision for every α. By (5.5), we can again subdivide $K^{(2)}_{2n}$ into a complex $K^{(3)}_{2n}$ of simplexes so that conditions (a) and (b) of §1 are satisfied with respect to T. Now, let ψ^α_{2s-1} and ϕ^α_{2s-1} be the chains obtained from $|\psi^\alpha_{2s-1}|$ and $|\phi^\alpha_{2s-1}|$ by taking the chain sum of their simplexes properly oriented. Since $F(X^*_{2s}) - F(\bar{X}^*_{2s}) = F(\Delta^{2s}_{2s}) = 0$, $F(X^*_{2s})$ is positively invariant. Let $F(X^*_{2s}) = \Delta^{2s}_{2s-1}$. We have $F(X^*_{2s}) = F\left(\sum_{\alpha<\beta} t_\alpha t_\beta E^\alpha_s \times E^\beta_s\right) + \Sigma\, t^2_\alpha \psi^\alpha_{2s-1}$.

[17] If s is even, ψ^α_{2s} is positively invariant and $\psi^\alpha_{2s} = \phi^\alpha_{2s} + \bar{\phi}^\alpha_{2s}$. The proof proceeds with no essential change.

Since ψ_{2s-1}^{α} is invariant and Δ_{2s-1}^{2s} is positively invariant, ψ_{2s-1}^{α} must be positively invariant. Thus we can write

$$\Delta_{2s-1}^{2s} = Y_{2s-1}^{*} + \bar{Y}_{2s-1}^{*} + \Sigma\, t_{\alpha}^{2}\, (\phi_{2s-1}^{\alpha} + \bar{\phi}_{2s-1}^{\alpha}) .$$

Let $X_{2s-1}^{*} = Y_{2s-1}^{*} + \Sigma\, t_{\alpha}^{2}\ \phi_{2s-1}^{\alpha}$, so that $\Delta_{2s-1}^{2s} = X_{2s-1}^{*} + \bar{X}_{2s-1}^{*}$. Since $\mathrm{Cl}\mid \phi_{2s-1}^{\alpha} \mid$ is the closure of a $(2s - 1)$-cell, $F(\phi_{2s-1}^{\alpha})$ is a simple circuit. Therefore all the $(2s - 2)$-simplexes of $F(\phi_{2s-1}^{\alpha})$ contained in S_{2s-2}^{α} occur in the symbol for $F(\phi_{2s-1}^{\alpha})$ with the coefficient one. Hence they occur in the symbol for $F(X_{2s-1}^{*})$ with the coefficient t_{α}^{2}. We continue this process until we have

$$\Delta_{s+1}^{2s} = X_{s+1}^{*} - \bar{X}_{s+1}^{*} .$$

The space S_{s}^{α}, defined by the equations $x_{q}^{\alpha} = y_{q}^{\alpha}$ $(q = 1, 2, \cdots, s)$, contains all the invariant points in $\mathrm{Cl}\mid \psi_{2s}^{\alpha} \mid$ and subdivides $\mid \psi_{s+1}^{\alpha} \mid$ by section into two sets $\mid \phi_{s+1}^{\alpha} \mid$ and $\mid \bar{\phi}_{s+1}^{\alpha} \mid$ of non-oriented convex cells which are interchanged by T and are such that $\mathrm{Cl}\mid \phi_{s+1}^{\alpha} \mid$ is the closure of an $(s + 1)$-cell. We introduce this subdivision for every α. By (5.5), we can further subdivide our complex into a complex $K_{2n}^{(l)}$ of simplexes which satisfies (a) and (b) of §1 with respect to T. Now, $F(\phi_{s+1}^{\alpha})$ is a simple circuit. Thus, in particular, all the cells $E_{s}^{0\alpha}$ (or the smaller simplexes by which they have been replaced during the successive subdivisions) of K_{n}^{0} which are on $F(\phi_{s+1}^{\alpha})$ occur in the symbol for $F(\phi_{s+1}^{\alpha})$ with the coefficient one. Hence they occur in the symbol for $F(X_{s+1}^{*})$ with the coefficient t_{α}^{2}. Thus we can write

$$F(X_{s+1}^{*}) = X_{s}^{*} + \bar{X}_{s}^{*} + \Sigma\, t_{\alpha}^{2}\, E_{s}^{0\alpha} = \Delta_{s}^{2s}, \text{ say} .$$

But $\Sigma\, t_{\alpha}^{2} E_{s}^{0\alpha} = \Sigma\, (t_{\alpha}^{2} - t_{\alpha}) E_{s}^{0\alpha} + \Sigma\, t_{\alpha} E_{s}^{0\alpha}$. Since $t_{\alpha}^{2} - t_{\alpha}$ is even, we can write $\Sigma\, t_{\alpha}^{2}\, E_{s}^{0\alpha} = U_{s}^{0} + \bar{U}_{s}^{0} + \Sigma\, t_{\alpha}\, E_{s}^{0\alpha}$. Let $W_{s} = X_{s}^{*} + U_{s}^{0}$ and $W_{s}^{0} = \Sigma\, t_{\alpha}\, E_{s}^{0\alpha}$. Then $\Delta_{s}^{2s} = W_{s} + \bar{W}_{s} + W_{s}^{0}$. The symbol for W_{s}^{0} has the same coefficients as that for a_{s}. Since K_{n} and K_{n}^{0} are identical in structure, the cycles W_{s}^{0i} form a base for weak homology of s-cycles of K_{n}^{0}.

(12.2) *The cycles* Δ_{h}^{2s} $(s = 0, 1, \cdots, n;\ h = s, s + 1, \cdots, 2s - 1)$ *of* (12.1) *satisfy the relation* $2\Delta_{h}^{2s} \simeq 0$.

For, $X_{h+1}^{*} \rightarrow \Delta_{h}^{2s}$ by definition, and $\bar{\Delta}_{h}^{2s} = \epsilon\, \Delta_{h}^{2s}$, $\mid \epsilon \mid = 1$. Therefore,

$$X_{h+1}^{*} + \epsilon\, \bar{X}_{h+1}^{*} \rightarrow 2\, \Delta_{h}^{2s} \simeq 0 .$$

Consider now a base for zero-divisors of dimension m of K_{2n}, like that of Lefschetz (loc. cit., p. 229) remembering that here $K_{2n} = K_{n} \times K_{n}$. Recalling the properties of our transformation T, it is easily seen that the cycles of this base can be renamed Ω^{i}, $\bar{\Omega}^{i} = T\Omega^{i}$, and Θ^{i}, where $T\Theta^{i} = \epsilon\Theta^{i}$, $\mid \epsilon \mid = 1$. The Θ's are the cycles of the form $b_{s}^{i} \times b_{s}^{i}$ $(m = 2s)$, where $c_{s+1}^{i} \rightarrow \zeta^{i}\, b_{s}^{i}$ on K_{n}.

(12.3) *Associated with each cycle* Θ_{2s}^{i} *of* K_{2n}, *there exists on* K_{2n}, *or a suitable subdivision of* K_{2n}, *a sequence of bounding cycles* $({}^{i}\Theta_{2s}^{2s}, {}^{i}\Theta_{2s-1}^{2s}, \cdots, {}^{i}\Theta_{s}^{2s})$ *and a sequence of chains* $({}^{i}\xi_{2s}^{*}, {}^{i}\xi_{2s-1}^{*}, \cdots, {}^{i}\xi_{s+1}^{*})$ *such that, for each value of* i, *we have*

$${}^{i}\Theta_{j}^{2s} = {}^{i}\xi_{j}^{*} + \epsilon\, {}^{i}\bar{\xi}_{j}^{*},\ \mid \epsilon \mid = 1;\ F({}^{i}\xi_{j}^{*}) = {}^{i}\Theta_{j-1}^{2s} \qquad (j = 2s, 2s - 1, \cdots, s + 1)$$

and ${}^{i}\Theta_s^{2s} \simeq \zeta^i Z_s^{0\,i}$ *on* K_n^0 *where the cycles* $Z_s^{0\,i}$ *constitute a base for zero-divisors of dimension* s *of* K_n^0.

Proof. We drop the superscript i temporarily and consider a definite $\Theta_{2s} = b_s \times b_s$ where $c_{s+1} \rightarrow \zeta\, b_s$. Consider the bounding relation

$$c_{s+1} \times c_{s+1} \rightarrow \zeta\, b_s \times c_{s+1} + (-1)^{s+1} \zeta\, c_{s+1} \times b_s.$$[18]

For the sake of definiteness let s be odd.[19] Then $c_{s+1} \times c_{s+1}$ is positively invariant and we can write $c_{s+1} \times c_{s+1} = \eta^*_{2\,s+2} + \bar{\eta}^*_{2\,s+2}$. We can choose $\eta^*_{2\,s+2}$ so that $\eta^*_{2\,s+2} \rightarrow \zeta\, b_s \times c_{s+1} - Y^*_{2\,s+1}$. Since

$$F(\eta^*_{2\,s+2}) + F(\bar{\eta}^*_{2\,s+2}) = F(c_{s+1} \times c_{s+1}),$$

we have

$$\bar{\eta}^*_{2\,s+2} \rightarrow \zeta\, c_{s+1} \times b_s + Y^*_{2\,s+1}.$$

Thus $Y^*_{2\,s+1}$ is negatively invariant and can be written as $\eta^*_{2\,s+1} - \eta^*_{2\,s+1}$. Now $F(Y^*_{2\,s+1}) = F(\zeta b_s \times c_{s+1}) = \zeta^2 b_s \times b_s = \Theta^{2\,s}_{2\,s}$, say. Evidently $\Theta^{2\,s}_{2\,s}$ is negatively invariant, and can be written as $\xi^*_{2\,s} - \bar{\xi}^*_{2\,s}$. Choose $\eta^*_{2\,s+1}$ so that $\eta^*_{2\,s+1} \rightarrow \xi^*_{2\,s} - Y^*_{2\,s}$. Since $F(\eta^*_{2\,s+1}) - F(\bar{\eta}^*_{2\,s+1}) = \xi^*_{2\,s} - \bar{\xi}^*_{2\,s}$, we have $\bar{\eta}^*_{2\,s+1} \rightarrow \bar{\xi}^*_{2\,s} - Y^*_{2\,s}$. Hence $Y^*_{2\,s}$ is positively invariant and can be written as $\eta^*_{2\,s} + \bar{\eta}^*_{2\,s}$. Now $F(Y^*_{2\,s}) = F(\xi^*_{2\,s}) = \Theta^{2\,s}_{2\,s-1}$, say. Since $\Theta^{2\,s}_{2\,s-1}$ is positively invariant, we write $\Theta^{2\,s}_{2\,s-1} = \xi^*_{2\,s-1} + \bar{\xi}^*_{2\,s-1}$. Choose $\eta^*_{2\,s}$ so that

$$\dot{\eta}^*_{2\,s} \rightarrow \xi^*_{2\,s-1} - Y^*_{2\,s-1}.$$

Since $F(\eta^*_{2\,s}) + F(\bar{\eta}^*_{2\,s}) = \xi^*_{2\,s-1} + \bar{\xi}^*_{2\,s-1}$, we have

$$\bar{\eta}^*_{2\,s} \rightarrow \bar{\xi}^*_{2\,s-1} + Y^*_{2\,s-1}.$$

Thus $Y^*_{2\,s-1}$ is negatively invariant. We proceed in this way until we have

$$Y^*_{s+2} = \eta^*_{s+2} - \bar{\eta}^*_{s+2} \rightarrow F(\xi^*_{s+2}) = \xi^*_{s+1} - \bar{\xi}^*_{s+1} = \Theta^{2\,s}_{s+1}, \text{ say,}$$

where η^*_{s+2} is chosen so that $\eta^*_{s+2} \rightarrow \xi^*_{s+1} - Y_{s+1}$. Now $\bar{\eta}^*_{s+2} \rightarrow \bar{\xi}^*_{s+1} - Y_{s+1}$. Thus Y_{s+1} is positively invariant. Finally we write $Y_{s+1} \rightarrow F(\xi^*_{s+1}) = \Theta^{2\,s}_s$, say. Let us trace the effect of this process upon a single convex cell $E^\alpha_{s+1} \times E^\alpha_{s+1}$ occurring with a non-zero coefficient in $c_{s+1} \times c_{s+1}$ where $c_{s+1} = \Sigma t_\alpha E^\alpha_{s+1}$. The totality of invariant points in $\mathrm{Cl}\,|\,E^\alpha_{s+1} \times E^\alpha_{s+1}\,|$ constitute the closure of an $(s+1)$-cell. Now, the separation of Y^*_h into $\eta^*_h + \epsilon\bar{\eta}^*_h$, $|\,\epsilon\,| = 1$, can be accomplished by subdivision by section by means of flat $(h-1)$ spaces S^α_{h-1}, as in (12.1). Now $\mathrm{Cl}\,|\,Y_{h-1}\,|$ is seen to be composed of points common to $\mathrm{Cl}\,|\,F(\eta^*_h)\,|$ and $\mathrm{Cl}\,|\,F(\bar{\eta}^*_h)\,|$. Hence $\mathrm{Cl}\,|\,Y_{s+1}\,|$ consists entirely of invariant points; therefore Y_{s+1} can be denoted by Y^0_{s+1}. Now, $Y^0_{s+1} = \Sigma t^2_\alpha E^{0\,\alpha}_{s+1}$, by the same argument as in (12.1). Then $Y^0_{s+1} = V^0_{s+1} + \bar{V}^0_{s+1} + \Sigma t_\alpha E^{0\,\alpha}_{s+1}$, as in (12.1). We have therefore $F(Y^0_{s+1}) = \Theta^{2\,s}_s = 2F(V^0_{s+1}) + F(\Sigma t_\alpha E^{0\,\alpha}_{s+1})$. But the chain symbol for $F(\Sigma t_\alpha E^{0\,\alpha}_{s+1})$ has the same coefficients as that for ζb_s, and can be de-

[18] Lefschetz, loc. cit., p. 227.

[19] If s is even, the proof is essentially the same.

noted by ζZ_s^0. Since K_n and K_n^0 are identical in structure, the cycles $Z_s^{0\,i}$ form a base for zero-divisors of dimension s of K_n^0. Evidently, each ${}^i\Theta_s^{2\,s} \simeq \zeta^i Z_s^{0\,i}$.

The cycles ${}^i\Delta_h^{2\,s}$ arising (12.1) from the cycles $\Delta_{2\,s}^i (s = 2n, 2n - 1, \cdots; i = 1, 2, \cdots, R_s(K_n))$ will be denoted generically by $\dot{\Delta}_h$. The cycles ${}^i\Theta_h^{2\,s}$ arising (12.3) from the cycles $\Theta_{2\,s}^i$ $(s = 2n, 2n - 1, \cdots; i = 1, 2, \cdots)$ will be denoted generically by $\dot{\Theta}_h$ except for $h = s$.

13. We shall need the following lemmas.

(13.1) *If $D + \epsilon\bar{D}$, $|\epsilon| = 1$, is a cycle, then $D + \epsilon\bar{D} \sim [\Gamma + \epsilon\bar{\Gamma}, \Delta, \Omega + \epsilon\bar{\Omega}, \Theta]'$.*

Proof. Let $\epsilon = 1$. We have

$$D + \bar{D} \sim \Sigma(a_i\Gamma^i + \bar{a}_i\bar{\Gamma}^i) + [\Delta]' + \Sigma(b_i\Omega^i + \bar{b}_i\bar{\Omega}^i) + [\Theta]'. \tag{13.2}$$

Applying T and subtracting the resulting homology from (13.2) we obtain

$$0 \sim \Sigma(a_i - \bar{a}_i)(\Gamma^i - \bar{\Gamma}^i) + [\Delta]' + \Sigma(b_i - \bar{b}_i)(\Omega^i - \bar{\Omega}^i) + [\Theta]'.$$

Hence $a_i = \bar{a}_i$. Also either $b_i = \bar{b}_i$ or $b_i - \bar{b}_i$ is a multiple of the order of Ω^i. In the latter case, we have $(b_i - \bar{b}_i)\bar{\Omega}^i \sim 0$, or $\bar{b}_i\bar{\Omega}^i \sim b_i\bar{\Omega}^i$. Thus in either case we can rewrite (13.2) as

$$D + \bar{D} \sim \Sigma a_i(\Gamma^i + \bar{\Gamma}^i) + [\Delta]' + \Sigma b_i(\Omega^i + \bar{\Omega}^i) + [\Theta]'.$$

This completes the proof. If $\epsilon = -1$, the proof is essentially the same.

(13.3) *If $D_m + \epsilon\bar{D}_m$, $|\epsilon| = 1$, is a cycle, then*

$$D_m + \epsilon\bar{D}_m \simeq [\Gamma_m + \epsilon\bar{\Gamma}_m, \Delta_m, \Omega_m + \epsilon\bar{\Omega}_m, \Theta_m]' + \{\dot{\Delta}_m\}' + [\dot{\Theta}_m]' + 2[Z_m^0]',$$

where no $\dot{\Delta}_m$ occurring with a non-zero coefficient is of the form $W_m + \bar{W}_m + W_m^0$.

Proof. By induction.

(A). By (13.1), the theorem is true for $m = 2n$, since for this highest dimension cycles of type $\dot{\Delta}$, $\dot{\Theta}$, and Z^0 do not exist, and $\simeq$ means only $=$.

(B). Assume the theorem for the dimension $m + 1 \leqq 2n$. Case I: let $\epsilon = 1$, $m = 2s$, s odd. By (13.1), there exists a chain H_{m+1} such that

$$H_{m+1} \rightarrow D_m + \bar{D}_m + [\Gamma_m + \bar{\Gamma}_m, \Delta_m, \Omega_m + \bar{\Omega}_m, \Theta_m]'. \tag{13.4}$$

Since s is odd, Δ_m and Θ_m are negatively invariant. Hence

$$H_{m+1} - \bar{H}_{m+1} \rightarrow 2[\Delta_m, \Theta_m]'.$$

But $2[\Delta_m, \Theta_m]'$ must be zero for otherwise we would have $2[\Delta_m, \Theta_m]' \sim 0$ which is impossible since the Δ's and Θ's are elements of a base for homology. Therefore $H_{m+1} - \bar{H}_{m+1}$ is a cycle, and by the induction hypothesis

$$\begin{aligned} H_{m+1} - \bar{H}_{m+1} \simeq [\Gamma_{m+1} - \bar{\Gamma}_{m+1}, \Omega_{m+1} - \bar{\Omega}_{m+1}]' + \{\dot{\Delta}_{m+1}\}' \\ + [\Theta_{m+1}]' + 2[Z_{m+1}^0]' \end{aligned} \tag{13.5}$$

since there are no cycles Δ_{2s+1} or Θ_{2s+1}. None of the cycles $\dot{\Delta}_{m+1}$ or $\dot{\Theta}_{m+1}$ oc-

curring with non-zero coefficients in (13.5) can be of the form $X + \bar{X}$, and $2[Z^0_{m+1}] = 0$, since the terms of (13.5) must be negatively invariant. Therefore $\dot{\Delta}_{m+1} = X_{m+1} - \bar{X}_{m+1}$ and $\dot{\Theta}_{m+1} = \xi_{m+1} - \bar{\xi}_{m+1}$. Now let

$$C_{m+1} = [\Gamma_{m+1}, \Omega_{m+1}]' + \{X_{m+1}\}' + [\xi_{m+1}]',$$

where the coefficients of the terms on the right are the same as in (13.5). Now (13.5) becomes $H_{m+1} - \bar{H}_{m+1} \simeq C_{m+1} - \bar{C}_{m+1}$, or

$$(H_{m+1} - C_{m+1}) - \overline{(H_{m+1} - C_{m+1})} \simeq 0 .$$

By (11.5), $F(H_{m+1} - C_{m+1}) \simeq 0$, or

$$F(H_{m+1}) \simeq F(C_{m+1}) .$$

By §12, $F(X_{m+1}) = \dot{\Delta}_m$ and $F(\xi_{m+1}) = \dot{\Theta}_m$ or ζZ^0_m. Thus,

(13.6) $D_m + \bar{D}_m + [\Gamma_m + \bar{\Gamma}_m, \Delta_m, \Omega_m + \bar{\Omega}_m, \Theta_m]' \simeq [\dot{\Delta}_m]' + [\dot{\Theta}_m]' + [Z^0_m]'$.

By (12.2), $[\dot{\Delta}_m]'$ can be replaced by $\{\dot{\Delta}_m\}'$. Therefore, no $\dot{\Delta}_m$ occurring in (13.6) with a non-zero coefficient can be of the form $W_m + \bar{W}_m + W^0_m$, for otherwise, remembering that the cycles $W^{0\,i}_m$ form a base for weak homology on K^0_n (12.1), we would have a contradiction of (11.1). There remains to be proved only that the expression $[Z^0_m]'$ in (13.6) is of the form $2[Z^0_m]'$. We can write (13.6) in the form $C_m + \bar{C}_m + [Z^0_m]' \simeq 0$. That is,

$$V^*_{m+1} + \bar{V}^*_{m+1} + V^0_{m+1} \rightarrow C_m + \bar{C}_m + [Z^0_m]' = C^*_m + \bar{C}^*_m + 2C^0_m + [Z^0_m]'.$$

Now let $V^*_{m+1} \rightarrow L^*_m + L^0_m$. Thus $V^*_{m+1} + \bar{V}^*_{m+1} \rightarrow L^*_m + \bar{L}^*_m + 2L^0_m$. Let $V^0_{m+1} \rightarrow J^0_m$. Now, $2L^0_m + J^0_m = 2C^0_m + [Z^0_m]'$. But $J^0_m \sim 0$ on K^0_n. Thus $2L^0_m - 2C^0_m \sim [Z^0_m]'$. Let $L^0_m - C^0_m = U^0_m$. Now

$$2U^0_m \sim [Z^0_m]' = A_1 Z^{0\,1}_m + A_2 Z^{0\,2}_m + \cdots .$$

But the Z^0_m's form a base for zero-divisors on K^0_n, by (12.3). Therefore the A's must be all even, for $U^0_m \sim \Sigma B_i Z^{0\,i}_m$ so that $2U^0_m \sim 2\Sigma B_i Z^{0\,i}_m$; hence $2B_i = A_i$. Thus $[Z^0_m]'$ is of the form $2[Z^0_m]'$. This completes the proof for Case I. The other cases (namely: $\epsilon = -1$, $m = 2s$, s odd; $\epsilon = \pm 1$, $m = 2s$, s even; and $\epsilon = \pm 1$, m odd) present no new difficulties and are proved in essentially the same way.

14. Let $\gamma^i_m = \Lambda\Gamma^i_m$. If $m = 2s$, we have $\Delta^i_{2s} = X^i_{2s} \pm \bar{X}^i_{2s}$ according as s is even or odd. Let $\delta^i_{2s} = \Lambda X^i_{2s}$ if s is even, and let $\delta^i_{2s} = 0$ if s is odd.

(14.1) *If s is even, δ^i_{2s} is a cycle and is not null.*

For $X^i_{2s} \rightarrow {}^i\Delta^{2s}_{2s-1} = {}^iX^*_{2s-1} - {}^i\bar{X}^*_{2s-1}$ by (12.1). Therefore

$$\Lambda X^i_{2s} \rightarrow \Lambda^i\Delta^{2s}_{2s-1} = 0.$$

Thus δ^i_{2s} is a cycle. Furthermore ΛX^i_{2s} is not null. For, suppose $\Lambda X^i_{2s} = 0$. Then $\Lambda'\Lambda X^i_{2s} = \Delta^i_{2s} = 0$, a contradiction.

(14.2) *The cycles* γ^i *and those cycles* δ^i *which are not null constitute a minimal base for weak homology of m-cycles of* k_{2n}.

Proof. The cycles γ^i, $\delta^i \neq 0$ are linearly independent with respect to homology. For, suppose there were a relation $\Sigma t_i\gamma^i + \Sigma s_i\delta^i \sim 0$. Applying Λ', we have $\Sigma t_i(\Gamma^i + \bar{\Gamma}^i) + \Sigma s_i\Delta^i \sim 0$ which contradicts the hypothesis that the cycles Γ^i, $\bar{\Gamma}^i$, Δ^i form a base on K_{2n}.

There remains to be proved that every m-cycle d of k_{2n} satisfies a relation of the form $d = [\gamma, \delta]'$. By (13.3),

$$\Lambda'd = D + \bar{D} \simeq [\Gamma + \bar{\Gamma}, \Delta, \Omega + \bar{\Omega}, \Theta]' + \{\dot{\Delta}\}' + [\dot{\Theta}]' + 2[Z^0]',$$

where Δ, $\dot{\Delta}$, Θ, $\dot{\Theta}$ can be written as $X + \bar{X}$, $x + \bar{x}$, $Y + \bar{Y}$, $y + \bar{y}$ respectively. Let

$$C = D + [\Gamma]' + [X]' + [\Omega]' + [Y]' + \{x\}' + [y]' + [Z^0]'.$$

Evidently $C + \bar{C} \simeq 0$. By (11.6),

$$(14.3)\quad d \sim [\gamma]' + [\delta]' + [\Lambda\Omega]' + [\Lambda Y]' + \{\Lambda x\}' + [\Lambda y]' + [\Lambda Z^0]' + z^0,$$

where z^0 is either null or a zero-divisor of k_n^0. It is easily seen that all the terms on the right of (14.3) are $= 0$ except the γ's and the δ's. Thus $d = [\gamma]' + [\delta]'$ which completes the proof.

Thus, $R_m(k_{2n}, 2^u)$, $u > 1$, is the number of cycles γ^i plus the number of cycles δ^i not null (14.1). Therefore,

THEOREM 6. *The numbers* $R_m(k_{2n}, 2^u)$, $u > 1$, *are given by formulae* (6.5) *where all the R's are now understood to be Betti numbers mod* 2^u.

BROOKLYN COLLEGE.

CONCERNING CERTAIN REDUCIBLE POLYNOMIALS

By H. L. Dorwart

1. **Introduction.** In a joint paper[1] by Oystein Ore and the author, it has been shown that polynomials with rational integral coefficients which take the values $\pm p$ (p a rational prime) for m integral values of the argument must take the same value $+p$ or $-p$ for $m > 5$, and consequently have the form

$$(x - a_1)(x - a_2) \cdots (x - a_m) h(x) \pm p.$$

In the same paper it has also been shown that integral polynomials of the form

$$f(x) = a(x - a_1) \cdots (x - a_n) \pm p \tag{1}$$

for $n > 6$ are irreducible in the rational domain if n is odd, and if n is even they may have only two factors of the degree $n/2$.[2] A new proof of this result is contained in a recent paper by A. Brauer.[3] In this paper, Brauer also raises the question whether or not these reducible polynomials can exist for every even n. A numerical example of eighth degree for the prime 2879 has been given by Pólya,[4] and one of tenth degree for the prime 10079 is contained in Brauer's paper. However, neither of these writers has obtained any general results on the subject.

We shall here find an expression for the necessary and sufficient conditions for the reducibility of these polynomials. From this result follows first the theorem of Brauer that the sum or difference of two such factors is a constant. Furthermore, it reduces the problem to well known problems in Diophantine equations for which partial solutions exist, and these solutions in turn give

Received by the Editors of the Annals of Mathematics February 27, 1934, accepted by them and later transferred to this journal. Presented to the American Mathematical Society, December 26, 1933.

[1] H. L. Dorwart and Oystein Ore, *Criteria for the irreducibility of polynomials*, Annals of Math., (2), vol. 34 (1933), pp. 81–94.

[2] That this theorem does not hold for $n = 6$ is shown by the following class of polynomials

$$x(x+1)(x-1)(x-2)(x-\alpha)(x-\beta) - p = [x^2 - x - 1][x^4 - 2x^3 - (p+1)x^2 + (p+2)x + p]$$

for primes of the form $\alpha^2 - (\alpha + 1)$, where α is a positive integer > 2 and $\beta = 1 - \alpha$.

[3] A. Brauer, *Bemerkungen zu einem Satze von Herrn G. Pólya*, Jahresber. Deutschen Math. Ver., vol. 43 (1933), pp. 124–129.

[4] Georg Pólya, *Verschiedene Bemerkungen zur Zahlentheorie*, Jahresber. Deutschen Math. Ver., vol. 28 (1919), p. 40.

various new decompositions for our polynomials. The new form of the problem also seems to indicate that decompositions exist for arbitrarily high degrees.

2. **Necessary and sufficient conditions.** Since we are considering only even degrees, let $n = 2m$. Next, let $k(x)$ be a factor of $f(x)$ which takes $+1$ for $a_1, \cdots, a_m$. For $m > 5$, it will have the form

$$k(x) = b(x - a_1) \cdots (x - a_m) + 1.$$

But $k(x)$ must take $\pm p$ for $a_{m+1}, \cdots, a_n$, and the conditions to be satisfied are

$$b(a_i - a_1) \cdots (a_i - a_m) = \pm p - 1 \qquad (i = m + 1, \cdots, n).$$

The other factor, say $l(x)$, must take $\pm p$ for $a_1, \cdots, a_m$, i.e.,

$$l(x) = c(x - a_1) \cdots (x - a_m) \pm p,$$

and must take ± 1 for $a_{m+1}, \cdots, a_n$. Hence the conditions

$$c(a_i - a_1) \cdots (a_i - a_m) = \pm 1 \mp p \qquad (i = m + 1, \cdots, n)$$

must also be satisfied. These conditions can exist simultaneously with those for $k(x)$ only if $b = c = \sqrt{a}$, and if the minus sign is chosen in front of the 1 and the signs in front of the p are reversed, i.e., $k(x)$ and $l(x)$ must take the values

	$a_1, \cdots, a_m$	$a_{m+1}, \cdots, a_n$
$k(x)$	$+1$	$+p(-p)$
$l(x)$	$-p(+p)$	-1

and the reducibility conditions are

$$\sqrt{a}(a_i - a_1) \cdots (a_i - a_m) = \pm p - 1, \qquad (i = m + 1, \cdots, n).$$

A similar situation results when it is assumed that $k(x)$ takes -1 for $a_1, \cdots, a_m$ etc. Hence we have the necessary and sufficient reducibility conditions for $m > 5$, i.e., $n > 10$. It can easily be shown that these conditions also hold for $n = 10$. For this degree, as stated in the introduction, decomposition is possible only in two factors of equal degree. Each factor must take ± 1 five times and $\pm p$ five times. However, it has already been shown[5] that it is not possible for such a factor to take five values ± 1 unless it takes only either $+1$ or -1 five times, in which case we have the situation discussed above. We can therefore say:

THEOREM 1. *The necessary and sufficient conditions for the reducibility of the polynomials* (1) *of even degree* $n = 2m$ *for* $n > 8$ *are*

$$\sqrt{a}(a_i - a_1) \cdots (a_i - a_m) = | p \pm 1 | \qquad (i = m + 1, \cdots, n), \tag{2}$$

[5] Dorwart and Ore, l.c., p. 85.

and the decomposition can be made in any one of the four equivalent forms

$$
(3)\quad
\begin{aligned}
&[\sqrt{a}(x-a_1)\cdots(x-a_m)\pm 1][\sqrt{a}(x-a_1)\cdots(x-a_m)\pm p]\\
&[\sqrt{a}(x-a_1)\cdots(x-a_m)\pm 1][\sqrt{a}(x-a_{m+1})\cdots(x-a_n)\mp 1]\\
&[\sqrt{a}(x-a_{m+1})\cdots(x-a_n)\mp p][\sqrt{a}(x-a_1)\cdots(x-a_m)\pm p]\\
&[\sqrt{a}(x-a_{m+1})\cdots(x-a_n)\mp p][\sqrt{a}(x-a_{m+1})\cdots(x-a_n)\mp 1],
\end{aligned}
$$

with proper attention to signs.

3. **Equivalent problems.** From Theorem 1, evidently a must be a perfect square and $\sqrt{a}$ must be a divisor of $|p \pm 1|$. Let $|p \pm 1| = \sqrt{a}\cdot d$. Then the conditions (2) reduce to

$$(4)\qquad (a_i - a_1)\cdots(a_i - a_m) = d \qquad (i = m+1, \cdots, n).$$

Let $\sigma_1, \sigma_2, \cdots, \sigma_m$ be the elementary symmetric functions of $a_1, \cdots, a_m$, and $\rho_1, \rho_2, \cdots, \rho_m$ be the elementary symmetric functions of $a_{m+1}, \cdots, a_n$. Then conditions (4) become

$$(5)\qquad a_i^m - \sigma_1 a_i^{m-1} + \sigma_2 a_i^{m-2} - \cdots + (-1)^m \sigma_m - d = 0 \qquad (i = m+1, \cdots, n),$$

which are easily seen to be equivalent to

$$(6)\qquad \sigma_1 = \rho_1, \quad \sigma_2 = \rho_2, \quad \cdots, \quad \sigma_{m-1} = \rho_{m-1}, \quad |\sigma_m - \rho_m| = d.$$

For a set of integers satisfying the first $m - 1$ of these conditions, the last one determines d, which in turn for a given $\sqrt{a}$ determines p. In fact, a well known theorem of Dirichlet[6] states that there will be an infinite number of primes p available.

A new wording of the first $m - 1$ conditions of (6) gives rise to

THEOREM 2. *The reducibility conditions of Theorem* 1 *are essentially equivalent to the problem of finding two equations of* m^{th} *degree differing only in the constant term, each having* m *distinct integral roots, and the* m *roots of the first equation all distinct from the* m *roots of the second equation.*

As Dickson has noted,[7] this problem came up in the early attempts to find rapidly converging series convenient for the computation of logarithms. A paper by E. B. Escott[8] contains most of the earlier results and shows how to compute numerical examples for $m = 3, \cdots, 7$.

4. Furthermore, it is known[7] that the problem of Theorem 2 is equivalent to a special case of the problem of Equal Sums of Like Powers of Diophantine analysis, i.e.,

THEOREM 3. *The reducibility conditions of Theorem* 1 *are essentially equivalent to the problem of finding the distinct integral solutions of the system*

[6] See Dickson, *History of the Theory of Numbers*, vol. 2, p. 415.
[7] L. c., footnote 6, p. 714.
[8] Quarterly Journal Math., vol. 41 (1910), pp. 141–167.

$$\sum_{i=1}^{m} a_i^k = \sum_{j=m+1}^{n} a_j^k \qquad (k = 1, \cdots, m-1).$$

General solutions of this problem have been found[9] for $m = 3, 4$ and numerous papers[10] give special solutions and numerical examples for $m = 5, \cdots, 8$. These in turn furnish reducible polynomials of the type (3) for $n = 8, 10, 12, 14, 16$. The following examples involve the smallest primes for degrees 8, 10, 12 that this writer has been able to find.

$$\begin{aligned}
&x(x-1)(x-2)(x-4)(x-7)(x-9)(x-10)(x-11) + 179\\
&\quad = [x(x-4)(x-7)(x-11) + 179][(x-1)(x-2)(x-9)(x-10) - 179].\\
&(x^2-1)(x^2-5^2)(x^2-7^2)(x^2-8^2)(x^2-9^2) + 5039\\
&\quad = [(x-1)(x-5)(x+7)(x+8)(x-9) + 5039]\\
&\qquad [(x+1)(x+5)(x-7)(x-8)(x+9) - 5039].\\
&(x^2-1)(x^2-5^2)(x^2-6^2)(x^2-9^2)(x^2-10^2)(x^2-11^2) + 100799\\
&\quad = [(x^2-1)(x^2-9^2)(x^2-10^2) - 100799]\\
&\qquad [(x^2-5^2)(x^2-6^2)(x^2-11^2) + 100799].
\end{aligned}$$

5. **Application of method.** In conclusion, we shall show that our methods can also be used to determine the integral polynomials of degree n which take $\pm$ the same integer N $2n$ times. They are necessarily of the form

$$a(x - a_1) \cdots (x - a_n) \pm N, \tag{7}$$

and the conditions to be satisfied are

$$a(a_i - a_1) \cdots (a_i - a_n) = \mp 2N \qquad (i = n+1, \cdots, 2n). \tag{8}$$

By the reasoning previously employed, these are equivalent to

$$\sigma_1 = \rho_1, \quad \sigma_2 = \rho_2, \quad \cdots, \quad \sigma_{n-1} = \rho_{n-1}, \quad |\sigma_n - \rho_n| = 2N/a \tag{9}$$

or

$$\begin{cases} \displaystyle\sum_{i=1}^{n} a_i^k = \sum_{j=n+1}^{2n} a_j^k & (k = 1, \cdots, n-1),\\ |\sigma_n - \rho_n| = 2N/a. \end{cases} \tag{10}$$

THEOREM 4. *Integral polynomials of degree n which take $\pm$ the same integer N $2n$ times can exist only for those values of N for which the conditions* (9) *or* (10) *can be satisfied.*

Examples of these polynomials are:

$$x(x-3) + 1 = (x-1)(x-2) - 1,$$

$$x(x-4)(x-5) - 6 = (x-1)(x-2)(x-6) + 6,$$

$$x(x-4)(x-7)(x-11) + 90 = (x-1)(x-2)(x-9)(x-10) - 90, \text{ etc.}$$

WILLIAMS COLLEGE.

[9] L. E. Dickson, *Introduction to the Theory of Numbers*, pp. 49–58.

[10] L. c., footnote 6, pp. 705–713.

ON SOME CHARACTERIZATIONS OF 2-DIMENSIONAL MANIFOLDS

By Egbertus R. van Kampen

1.1. **Object.** A large number of papers have been devoted to the problem of finding topological characterizations for 2-sphere, 2-cell or 2-dimensional manifolds (finite or infinite) of different type. Through complicated cross-citation on the one side, improvements in the available methods on the other side, the study of these papers seems to be at present so much harder than necessary for anybody not already thoroughly acquainted with the ideas used, that the publication of a systematic, simplified exposé of the attained results seems to be the only way of giving these results the place they deserve in the theory of point sets. The results could of course be simplified and extended in several directions. In an appendix we prove that the 2-dimensional generalized manifolds of Čech and Lefschetz are ordinary manifolds.

1.2. **Outline of contents and methods.** The Theorems I to V″ of this paper are very closely related in formulation and proof. This formulation can be reduced to the following scheme. A compact or locally compact Peano space contains at least one curve of one of a few simple types; every curve of that type separates and no closed subset of such a curve separates the space; then the space is homeomorphic with some type of 2-dimensional manifold. In this way we treat in I and III the 2-sphere, in II the closed 2-cell, in IV the 2-dimensional manifold without boundary, in V the open (infinite) 2-dimensional manifolds. The investigation of the set of conditions in IV for a 2-dimensional manifold without boundary was suggested by Zippin. We could have given a characterization of the 2-dimensional manifolds with boundaries by suitably combining the conditions of II and IV. As the result is less elegant and its formulation and proof do not need any additional idea, we leave this to the reader.

The proofs show of course the effect of the similarity in statement. In later proofs many arguments have been left out simply because they have already occurred before. In the proof of I the most important part is the cutting up of the space by a linear graph in arbitrarily small pieces. Different ways of doing this have been used by Moore, Gawehn, Radó (3). We finally used directly a method of approximating a sum of arcs by a linear graph suggested originally by Zippin for the proof of a lemma. The whole argument has been formulated in such a way that it can be used without any change several times more.

Theorems VI and VII contain characterizations by means of Vietoris chains of the compact types of 2-dimensional manifolds. Theorem VII was proved by

Received by the Editors of the Annals of Mathematics February 10, 1934, accepted by them, and later transferred to this journal.

H. Whitney (14). The theorems are very easily reduced to preceding theorems. We suppose knowledge of the combinatorial notions and theorems used.

In the rest of the paper we give an account of the consequences that can be drawn for Peano spaces from different forms of various parts of the theorem of Janiszewski. Some of the theorems given were proved by Kuratowski (8) and Zippin (9). The proofs in that section are held rather short, but they are all of classical type so that more elaboration seemed superfluous.

1.3. **Review of the literature.**[1] In (1) Moore gave three systems of axioms for plane topology, proving in (2) that the spaces determined by his systems were really homeomorphic with the plane. In (3) Moore and Kline announce that certain conditions, of which the important ones are part of or closely related to the Jordan curve theorem, are sufficient to characterize the 2-sphere among point sets situated in 3-space. Their proof does not seem to have been published. In (5) Miss Gawehn proves independently that a slightly modified set of conditions will define 2-dimensional manifolds without boundary among arbitrary Hausdorff spaces.

With (6) the line of attack undergoes a definite change. There Kuratowski announces a system of axioms for spherical topology of which the principal part is formed by the theorem of Janiszewski. His proofs will be found in (8). In the system of axioms given by Woodard in (7) following more closely the scheme of Moore in (1) the most important axiom shows some similarity to the Janiszewski theorem too. In (9) Zippin investigates what kind of Peano continua satisfy the Janiszewski theorem. His results are that in compact (locally compact) locally connected, connected spaces satisfying the Janiszewski theorem, the non-degenerate cyclic elements are homeomorphic with a 2-sphere (a region on a 2-sphere). A systematic account of the results in this line will be found from 9.1 on in this paper.

Again a change takes place. In (10) Wilder announces and in (11) he proves a characterization of the sphere for the purely utilitarian purpose of characterizing the domains determined in 3-space by a sphere. He uses again elements of or related to the Jordan curve theorem. Attacking the problem on its own merits, Zippin then finds in (12) considerably simpler results capable of extension to regions on the sphere. He treats the subject in two different ways (pp. 333–340 and pp. 341–349 of (12)). The results he arrives at by the second treatment cannot be considered as final and are not considered in this paper. The results of the first way of treatment will be found in the theorems I, I′ and V′ of this paper. In (15) Zippin gives a characterization of a closed 2-cell closely analogous to the first treatment of the sphere in (12). His result will be found in Theorem II of this paper. In the meantime Roberts (13) had given a characterization of 2-dimensional manifolds without boundary and Whitney (14) had given a characterization of the closed 2-cell using an entirely different and

[1] Numbers in parentheses refer to the list of literature at the end of this paper. Less common terms will be explained in sections 2.1 to 2.4.

very powerful condition (namely the existence of a certain type of chain). See for his result theorem VII of this paper.

2.1. All spaces mentioned in this paper will be separable, metric spaces. We call a space *compact*, if every infinite number of points in it has a limit point in it; *locally compact*, if each point has a neighborhood with compact closure; *connected*, if the space is not the sum of two closed and open subsets that have no point in common; *locally connected*, if each point is in arbitrarily small connected open sets.

2.2. A locally compact, connected and locally connected space we call *P-space* (Peano).

A P-space is arcwise connected. Any connected open subset of a P-space is again a P-space. If a P-space H is closed in a P-space K, the sum of H and any number of components of $K-H$ is again a P-space.

2.3. A P-space is called *cyclic* if it has no cut points, that is, if it is not disconnected by the removal of one point. Any two points on a cyclic P-space K are on a simple closed curve in K; any three points on K are in arbitrary order on an arc in K.

If a non-degenerate point set H in a P-space K is not separated on K by any cut point of K, then H determines uniquely a maximal cyclic subset of K, the *cyclic element* determined by H in K. Such a cyclic element C is a P-space, closed in K, contains all arcs in K with end points in C and each point of $K - C$ can be separated from C by a cut point on C.[2]

2.4. If an arc is not determined by a small Greek letter, we may name it by writing down in the correct order all letters representing points on that arc that have been named. If α or pqr is a closed arc, $< \alpha >$ or $< pqr >$ is the corresponding open arc.

We say that an arc spans a point set if it has the two end points and nothing else in common with that point set.

3.1. THEOREM I. *A compact P-space D satisfying the following three conditions is homeomorphic with a 2-sphere:*

Ia. *D contains at least one simple closed curve.*

Ib. *Every simple closed curve of D separates D.*

Ic. *No closed arc of a simple closed curve of D separates D.*[3]

The proof of this theorem has been so constructed that the last part (4.1) to (4.5) can be used without any change in the proof of later theorems.

3.2. The following condition is equivalent to Ic:

[2] We consider here only non-degenerate cyclic elements. For proofs of the theorems mentioned see G. T. Whyburn, *On the cyclic connectivity theorem*, Bull. Am. Math. Soc., vol. 37 (1931), pp. 429–433. We have substituted the simpler term cyclic for cyclically connected.

[3] Ia and Ic could be replaced by the slightly stronger condition: *No closed arc of D separates D.* This can be proved by the methods used for Id and If.

Id. *Every component of the complement of a simple closed curve α of D has all points of α as limit points.*

It is trivial that Ic follows from Id. The converse is proved as follows. If any component of $D - \alpha$ did not have all points of α as limit points, it would be separated from the rest of α by a closed arc containing all those limit points.

Ie. *D is cyclic.*

This follows from Ia and Ic and implies of course Ia. It is equivalent to the statement that the cyclic element (see 2.3) of D containing a certain simple closed curve is equal to D. If this were not true, the cyclic element would contain a cut point of D through which passes a simple closed curve, in contradiction with Ic.

3.3. Take a simple closed curve J in D. *D cannot contain three arcs λ_1, λ_2 and μ spanning J with the following properties: $< \lambda_1 >$ and $< \lambda_2 >$ are in different components of $D - J$. μ does not meet λ_i and its end points are separated on J by those of λ_i, $(i = 1, 2)$.*[4]

$J + \lambda_1 + \lambda_2$ contains a simple closed curve J' not meeting μ, such that $J + \lambda_1 + \lambda_2 + \mu - J'$ consists of two open arcs in J connected by μ. Each component A of $D - J'$ must contain points of the components B_1 and B_2 of $D - J$ containing $< \lambda_1 >$ and $< \lambda_2 >$ (see Id). As B_1 and B_2 are different, A must contain points of J. But all the points of J left over are on one component of $D - J'$. This gives a contradiction to Ib.

If. *Every simple closed curve of D separates D into exactly two components.*

In any component A of $D - J$ we can draw a spanning arc of J of which the end points are on two given open arcs of J. For A plus the open arcs is a connected set because of Id. It is a P-space for it is a connected open subset of the closure of A (2.2). Accordingly it is arcwise connected and If follows because we could now draw three arcs like λ_1, λ_2 and μ if we had three components available.

If can of course replace Ib. If + Id is the Jordan curve theorem. Ia + If + Id is equivalent to Ia + Ib + Ic and so Theorem I is equivalent to the following:

THEOREM I'. *A compact P-space satisfying non-vacuously the Jordan curve theorem is homeomorphic with a 2-sphere.*

3.4. *If α, β and γ are three arcs with common end points, then $D - (\alpha + \beta + \gamma)$ is the sum of three components having $\alpha + \beta$, $\beta + \gamma$, $\alpha + \gamma$ as their boundaries.* The component separated by $\alpha + \beta$ from $< \gamma >$ and two analogous ones are separated in $D - (\alpha + \beta + \gamma)$ and have the correct boundaries. Any additional component E would have to possess boundary points in the interior of α, β and γ for otherwise one of the three simple closed curves would separate D into three components. The contradiction of 3.3 can now be found using $\alpha + \beta$ as simple closed curve, an arc in E from a point of $< \alpha >$ to a point of

[4] The addition of simple drawings here and at several other points might have simplified the presentation. However, nobody who is sufficiently interested will have any trouble supplying them himself.

$< \beta >$ as μ, γ as λ_1, and a spanning arc of $\alpha + \beta$ in the component with the boundary $\alpha + \beta$ as λ_2.

4.1. A simple closed curve J divides D into two components. We will prove that the closure C of any one component is now homeomorphic with a 2-cell, such that J corresponds to the boundary of the 2-cell. Theorem I is then a trivial consequence.

A subdivision of C by means of a graph G is called a *C-complex* $[C]$ if the graph $G = J + \sum_{i \leq n} \alpha_i$ can be constructed by doing n times the following step: add to $J + \sum_{i<m} \alpha_i$ a spanning arc α_m of $J + \sum_{i<m} \alpha_i$ in C. The 0-, 1-, and 2-cells of $[C]$ are the 0- and 1-cells of G and the components of $C - G$. The boundary of a 2-cell consists of all 1-cells in the interior of which the 2-cell has a limit point.

If a C-complex $[C]$ is given, we can find a subdivision of the 2-simplex C' by means of polygons such that the resulting complex $[C']$ is isomorphic with $[C]$.

The word *isomorphic* means that a one-to-one correspondence between the cells of the two complexes can be established such that the relation *on the boundary of* is invariant under that correspondence.

This can be proved by induction on the number of spanning arcs used in constructing the graph G defining $[C]$. We assume the theorem is proved for G_1, defining $[C]_1$, and we will prove it for $G_2 = G_1 + \alpha$, defining $[C]_2$, where α is an arc in C spanning G_1. The 2-cell A of $[C]_1$ containing $< \alpha >$ has for its boundary a simple closed curve, divided by α into two arcs β_1 and β_2. Corresponding to these we have in $[C']_1$ the arcs β_1' and β_2' and a 2-cell A'. In A' we can construct a polygonal 1-cell α' joining the end points of β_1' (or β_2'). This divides A' into two 2-cells B_i' with boundaries $\beta_i' + \alpha'$, $(i = 1, 2)$. But, according to 3.4, A is divided by α into two components with boundaries $\alpha + \beta_1$, and $\alpha + \beta_2$, so that the isomorphism has been extended to $[C_2]$.

If a subdivision of C' into a C'-complex $[C']$ is given, we can find by the same method an isomorphic C-complex $[C]$ provided that each spanning arc used in the construction of $[C']$ has its end points in the interiors of two 1-cells of the complex at that stage of construction. If we did not make this restriction, we would have to prove accessibility of points of simple closed curves in C in order to find the arc corresponding to some arc in C'. It is possible to do this, but entirely superfluous.

The closure of each component of $D - G$ is cyclic. See for the proof Ie in 3.2.

4.2. *A pair of points p and q in C can be separated by a spanning arc of J.* If p is not in J, we can find a spanning arc of J through p. If it meets q, we can easily change it so that it does not meet q. Then if q is not in J, we can find a spanning arc through q and change it so that it does not meet the spanning arc through p. Now p and q are on a simple closed curve containing two subarcs of J separating p and q. A spanning arc in C of that simple closed curve between points of those subarcs will separate p and q.

We now construct a finite number of arcs, not necessarily forming a graph, and dividing C into arbitrarily small pieces. For a given $\delta > 0$ we cover D by a finite

number of closed sets V_i, $i = 1, \cdots, n$ of diameter less than $\delta/2$. Because V_i is compact we can find for any point p of $C - V_i$ a finite set of spanning arcs of J such that p is separated from any point of V_i by at least one of those arcs. The set of points p separated from any point of V_i by one of those arcs contains a neighborhood of p, so that we can find a finite set of such finite sets of arcs, such that if $V_i \cdot V_j = 0$, any point p of V_i is separated from any point q of V_j by one of those arcs. Doing this for all pairs of sets V_i having no points in common, we find a finite set of spanning arcs of J, $\alpha_1 \cdots \alpha_p$, such that any pair of points, taken out of any pair of sets V_i having no point in common, is separated by one of those arcs.

4.3. *We construct a sequence of C-complexes $[C]_i$ defined by graphs G_i contained in $J + \Sigma\alpha_i$ and approximating $J + \Sigma\alpha_i$.* They will depend on a sequence of numbers ϵ_i, $\epsilon_i > \epsilon_{i+1} > 0$, $\epsilon_i \to 0$. G_0 is equal to J and equal to G_0^0. If the graph G_i^j has been defined and $j < p$, we take for G_i^{j+1} the sum of G_i^j and all sub-arcs of α_{j+1}, spanning G_i^j, such that the end points of each sub-arc cannot be joined in G_i^j by an arc of diameter less than ϵ_i (for instance, $G_0^1 = J + \alpha_1$ if ϵ_1 is not too big). The number of such arcs is finite because G_i^j is locally connected, so that G_i^{j+1} is really a graph. If G_i^j has been defined and $j = p$, we put $G_i^p = G_i = G_{i+1}^0$.

The graphs G_i have the property that *for a certain m, G_m contains a spanning arc of J separating any pair of points taken from any pair of sets V_i having no point in common.* If this were not true, we could find a sequence of such pairs of points p_i, q_i such that G_i contains no arc separating p_i and q_i. For some sequence of integers ν_i, p_{ν_i} and q_{ν_i} will both converge. We call the limits p and q. They will be in some sets V_a and V_b having no point in common, so that there will be an arc α_l separating them, with end points x and y. No G_i will contain an arc separating them. For any such arc would separate p_{ν_j} and q_{ν_j} for some $\nu_j > i$ and be contained in G_{ν_j}, contrary to our assumption.

We join p and q to J by two arcs γ_1 and γ_2 not meeting α_l and determine a number k such that ϵ_k is less than the distance from $\gamma_1 + \gamma_2$ to α_l. If U is an ϵ_k neighborhood of α_l in C, then x and y must be in one connected set of $U' = U \cdot G_k$. For if $U_x + U_y$ is a separation of U' between x and y, U_x will contain a last point r of α_l and U_y will contain a first point s after r of α_l. Then G_k must contain an arc of diameter less than ϵ_k between r and s and this arc must be in U', which is absurd. Then U' must contain an arc $\beta = xy$ and β must contain a spanning arc of J separating the end points on J of γ_1 and γ_2, and accordingly separating p and q.

4.4. *The C-complex $[C]_m$ defined by G_m is of mesh less than δ,* for any two points on the closure of the same 2-cell of $[C]_m$ must be on the same or adjoining sets V_i, so that their distance must be less than δ.

We construct a polygonal subdivision $[C']_m$ of C' similar to $[C]_m$ and a polygonal subdivision $[C']_\delta$ of $[C']_m$ (taking care that any new end point is in the interior of some old 1-cell) of mesh less than δ, finally a subdivision $[C]_\delta$ of $[C]_m$ similar to $[C']_\delta$.

We have seen how to construct for C and C' similar subdivisions $[C]_\delta$ and $[C']_\delta$ of

mesh less than δ. But then we can construct a sequence of such pairs of similar subdivisions $[C]_{\delta_i}$ and $[C']_{\delta_i}$, where $\delta_i \to 0$, for we can repeat the process applied to C, for the closure of any 2-cell of $[C]_{\delta_i}$.

4.5. The sets of 0-cells of all $[C]_{\delta_i}$'s and $[C']_{\delta_i}$'s form *two one-to-one corresponding everywhere dense point sets in C and C'. This correspondence is uniformly continuous both ways.* Let $\epsilon > 0$ be given; find n such that $2\delta_n < \epsilon$, now find $\delta > 0$ such that two points of distance less than δ in C or in C' are in or on the boundary of the same or adjoining 2-cells of $[C]_{\delta_n}$ or $[C']_{\delta_n}$. If such a δ did not exist, we could find two sequences of points p_i and q_i convergent to the same point p of C (or C') and not in adjoining 2-cells. We can assume that all p_i and all q_i are on the closure of the same 2-cell of $[C]_{\delta_n}$ (or $[C']_{\delta_n}$) because there are only a finite number of such 2-cells to choose from. But the point p belongs to the closure of both 2-cells so that they were adjoining after all.

If two vertices of any subdivision of C (or C') have a distance less than δ, their corresponding vertices in C' (or C) are in or on the boundary of the same or adjacent 2-cells of $[C']_{\delta_n}$ (or $[C]_{\delta_n}$), so that their distance is less than ϵ. This makes the correspondence uniformly continuous.

Now the correspondence of the two sets of vertices can be extended to a topological transformation of their closures. But these closures are identical with C and C', so that C and C' are homeomorphic. This can of course be extended immediately to prove that D is homeomorphic with the 2-sphere D'.

5.1. THEOREM II. *A compact P-space C containing a simple closed curve J and satisfying the following three conditions is homeomorphic with a closed 2-cell:*

IIa. *C contains an arc spanning J.*

IIb. *Every arc of C spanning J separates C.*

IIc. *No proper closed subset of an arc spanning J separates C.*[5]

Section 5.2 is closely analogous to 3.2 and 3.3. In 5.3 and 5.4 we prove that each component of the complement of a spanning arc of C contains a point of J. In 5.5 we prove that $C - J$ can be considered connected, which makes C cyclic and in 5.6 we prove that each part of C determined by a spanning arc has all the properties we supposed for C. After this point the proof of Theorem II is identical with the sections 4.1 to 4.5 of the proof of Theorem I, and may be omitted.

5.2. Let α be an arc spanning J. *Every component of $C - \alpha$ has every point of α as a limit point* (see 3.2).

C cannot contain three arcs λ_1, λ_2 and μ with the following properties: λ_1 and λ_2 span $J + \alpha$ and their interiors are contained in different components of $C - \alpha$. μ spans α and does not meet λ_i. One end point of λ_i is on α between the end points of μ and separated on $J + \alpha$ by the end points of μ from the other end point of λ_i. These other end points of λ_i are separated on α by each end point of μ if they are on α.

[5] Here as analogously in Theorem I conditions IIa and IIc can be replaced by the following slightly stronger one: C is not disconnected by the sum of two arcs that do not meet and each of which has one end point in common with J.

If such arcs λ_1, λ_2 and μ existed, we could find an arc δ spanning J, consisting of λ_1, λ_2 and 0, 1, 2 or 3 arcs on α, depending on whether the first end points of λ_i mentioned coincide or not and on whether the other end points are on $< \alpha >$ or not. Every component of $C - \delta$ would have to contain points of α (see 3.3), but all remaining points of α are on one component of $C - \delta$, so that $C - \delta$ would be connected. It follows that *$C - \alpha$ consists of exactly two components* (see If in 3.3).

5.3. *From now on till a contradiction is proved we will suppose that $C - \alpha$ has a component A, not containing a point of J.* Let the end points of α be called x and y.

$A' = \bar{A} - x$ is cyclic.[5a] $< \alpha >$ does not contain a cut point of A' because we can draw an arc in A' having its end points on two given open intervals of α (see If in 3.3). So α must belong to a cyclic element of A'. If this is not equal to A', it contains a cut point z of A' through which passes an arc xzy. But then the subset $x + z$ of xzy would disconnect C in contradiction with IIc.

We will say that an *arc β covers t* if β has end points x and y, contains a subarc xz of α, is contained in $\bar{A}$, and if t is contained in the component of $C - \beta$ containing $C - \bar{A}$. *Each point t of A' is covered by at least one arc.* Because A' is cyclic, it contains an arc zty spanning $< \alpha > + y$. In the component of $C - xzty$ not containing $C - \bar{A}$ we can construct an arc $< sry >$, s on $< zt >$. Then $xzsry$ covers t.

If U is an open subset of $\bar{A}$ containing x and $\bar{U}$ does not contain y, then each point of the (compact) boundary A'' of U can be covered by some arc, and each arc covers an open subset of A''. Accordingly we can find a finite set of arcs $\alpha_1 \cdots \alpha_p$ covering all points of A''. In the next section we will prove a contradiction starting from this statement.

5.4. We define *a sequence of graphs G_i approximating $< \alpha > + \Sigma\alpha_i$* analogous to the sequence of 4.3. *There must be an m such that for each point p of A'' there is an arc contained in G_m covering p.* Otherwise we could find by the same method as used in 4.3 a point p in A'' covered by α_l but not covered by any arc contained in any G_i. Join p to a point of α by an arc γ not meeting α_l. Then some G_i must contain an arc xy not meeting γ and that arc must cover p.

We prove that G_i contains an arc β_i covering all points of $G_i - \beta_i$. Remembering how G_i was constructed, we only need to prove that if any graph G contained in $\alpha + \Sigma\, \alpha_i$ contains an arc β covering $G - \beta$, and if γ is an arc in $\alpha + \Sigma\, \alpha_i$ spanning G, then $G' = G + \gamma$ contains an arc β' covering $G' - \beta'$. If β covers $< \gamma >$, we may put $\beta' = \beta$. If not, $< \gamma >$ is contained in the component of $C - \beta$ not meeting J and it must span β. Its end points determine a subarc γ' of β and we may put $\beta' = \beta + < \gamma > - < \gamma' >$.

The arc β_m of G_m covers all points covered by any other arc β' of G_m. If B_m (B') is the component of $C - \beta_m$ ($C - \beta'$) not meeting J we must prove $B_m \supset B'$. Because $C - B_m \subset \beta'$, the only alternative is $B_m \cdot B' = 0$. The arcs β_m

[5a] The symbol $\bar{A}$ is used to denote the closure of A.

and β' must have some subarc xz of α in common because all original arcs α_i must have such a subarc in common. Along xz we would have three point sets: B_m, B' and $C - \bar{A}$ in each of which we could draw an arc with end points on any two open subarcs of xz. This is impossible according to 5.2.

The arc β_m would now have to cover all points of A'', and this is impossible because β_m must contain points of A''. The statement, of which the proof is now complete, can be formulated thus: C cannot contain two arcs which span J, which have no point in common and whose pairs of end points are separated on J.

5.5. *Every point of J is a limit point of any component M of $C - J$ which contains a spanning arc of J.* If not, then there exists a maximal subarc $< \beta > = < pq >$ of J such that no point of it is a limit point of M ($\beta \neq J$ because M contains a spanning arc of J). $M + J - < \beta >$ contains a cyclic element $N \supset p + q$, for if $J - \beta$ contained a cut point of $M + J - < \beta >$, M would not be connected. The simple closed curve ϵ in N through p and q contains an arc $\gamma = pq$ spanning J. For if it contains an arc $< \delta > = < pq >$ meeting J, δ plus the components not containing $< \beta >$ of C minus any subarc of δ spanning J contains $J - \beta$, so that $\gamma = \epsilon - < \delta >$ must span J. Now $p + q$ would be a subset of a spanning arc separating C which is impossible.

It follows immediately that every component of $C - J$ different from M has exactly one limit point on J. If we take these components away, C will satisfy the additional condition that $C - J$ is connected. We will prove that C is homeomorphic with a 2-cell under this condition, and then it follows immediately that no components except M can have formed part of $C - J$. Under the additional condition C *is cyclic* (see Ie in 3.2).

5.6. *If A is a component of $C - \alpha$ and $< \lambda >$ the part of J it contains, then $\bar{A}$ has all the properties we assumed for C with $\alpha + \lambda$ as exceptional simple closed curve.* That $\bar{A}$ contains a spanning arc of $\alpha + \lambda$ and is cyclic is easily proved. We have to prove:

Ib. *If β is an arc in $\bar{A}$ spanning $\alpha + \lambda$, then $\bar{A} - \beta$ is not connected.* If both end points of β are on J the separation of C by β defines immediately a separation of $\bar{A}$ by β. So we assume that at least one end point of β is on α. There is an arc γ spanning J, contained in $\alpha + \beta$ and containing β. The part of α not on γ is a subarc of α that we call δ. We assume that β does not separate $\bar{A}$. Then we can find an arc pq in $\bar{A}$ from a point p on $< \delta >$ to a point q of $\alpha - \delta$. The arc β, the arc pq and an arc in $C - \bar{A}$ from a point on α between p and q to a point of J are three arcs in the contradictory position of 5.2, so that Ib follows for $\bar{A}$.

Ic. *If ϵ is a subarc of an arc β in $\bar{A}$ spanning $\alpha + \lambda$, then $\beta - < \epsilon >$ does not separate $\bar{A}$.* Defining γ and δ as before, we can join any two points of $\bar{A} - \gamma$ by an arc in $C - (\gamma - < \epsilon >)$. All parts of that arc in $C - \bar{A}$ can be replaced by subarcs of δ, so that the two points can be joined by an arc in $\bar{A} - (\beta - < \epsilon >)$.

From this point on, the proof of Theorem II is identical with that of Theorem I from 4.1 on, and may be omitted.

6. THEOREM III. *If we restrict all simple closed curves mentioned in the formulation of Theorem I to pass through a fixed point p of D, D is still homeomorphic with a 2-sphere.*

Id will still be true for simple closed curves containing p. Ie can still be proved in the same way as before. The contradiction at the beginning of 3.3 will still hold if we force J' to pass through p by the additional condition that p is on $\lambda_1 + \lambda_2$, or p is separated on J by the end points of λ_1 and λ_2 from the end points of μ. If is still true for simple closed curves on p. It is now readily possible to free conditions Ib and Ic from the restriction that the simple closed curve must pass through p.

Ib. Let J be a simple closed curve on D not passing through p. Draw a spanning arc γ of J through p dividing J into arcs α and β. Call the components of $D - (\alpha + \gamma)$: $A_1 \supset < \beta >$ and A_2; the components of $D - (\beta + \gamma)$: $B_1 \supset < \alpha >$ and B_2; then $B_2 \subset A_1$ and $A_2 \subset B_1$, so $A_2 \cdot B_2 = 0$ and $A_1 + B_1 = D - \gamma$. $A_1 \cdot B_1$ and $A_2 + B_2$ are separated parts of $D - (\alpha + \beta) - < \gamma >$. We will prove that $A_1 \cdot B_1$ and $A_2 + B_2 + < \gamma >$ are separated parts of $D - (\alpha + \beta)$. Obviously $< \gamma >$ has no limit points on A_1 or B_1, so we only have to prove that $A_1 \cdot B_1$ has no limit points on $< \gamma >$. If a component A_3 of $A_1 \cdot B_1$ had a limit point on $< \gamma >$, it would have to have limit points on $< \alpha >$ and $< \beta >$ because otherwise $\alpha + \gamma$ or $\beta + \gamma$ would separate D into at least 3 components. Now a graph with the contradictory properties of 3.3 could be constructed. The simple closed curve is $\alpha + \gamma$. For λ_1 we can take an arc in $A_3 \subset A_1$ from a point of $< \gamma >$ to a point of α; for λ_2 an arc in A_2 from a point of γ to a point of α such that p is separated on $\alpha + \gamma$ by the end points of λ_1 and λ_2 from the end points of β. For the arc μ we can take β.

Ic. The component containing p $(A_2 + B_2 + < \gamma >)$ determined by a simple closed curve J in D has every point of J as a limit point. So we can draw a spanning arc of J through p and points in two given open subarcs of J. If we want to prove that the closed subarc α of J does not separate D, we choose these open arcs on $J - \alpha$. Then α turns out to be a subarc of a simple closed curve through p and cannot separate D.

7.1. THEOREM IV. *A compact P-space B satisfying for a certain number $\epsilon > 0$ the following three conditions is homeomorphic with a 2-dimensional manifold without boundary.*

IVa. *B contains at least one simple closed curve of diameter less than ϵ.*

IVb. *Every simple closed curve of diameter less than ϵ in B separates B.*

IVc. *No closed arc of a simple closed curve of diameter less than ϵ in B separates B.*

THEOREM V. *If B is a P-space, such that for every compact subset F of B there is a number $\epsilon_F > 0$, for which the following three conditions are satisfied, then B is homeomorphic with a 2-dimensional infinite manifold.*

Va. *B contains a simple closed curve J of diameter $< \epsilon_J$.*

Vb. *A simple closed curve in F of diameter $< \epsilon_F$ separates B.*

Vc. *No closed subarc of a simple closed curve in F of diameter $< \epsilon_F$ separates B.*

The almost identical proofs of these theorems will be given simultaneously.

7.2. *For each non-cut point p of B and sufficiently small $\delta > 0$, we can find a number $\eta(\delta) > 0$, such that all points of which the distance to p is $\geqq \delta$, are in one component of the set of points of which the distance to p is $\geqq \eta(\delta)$.* Otherwise we could find two converging sequences of points $q_i \to q$ and $r_i \to r$ and a sequence of numbers η_i converging to zero, such that the distances $p - q_i$ and $p - r_i$ are $\geqq \delta$ and that any arc joining q_i and r_i comes within a distance η_i of p. This contradicts the fact that p and q can be joined by an arc not meeting p.

7.3. *A simple closed curve β within a distance η $(\epsilon/2)$ of a non-cut point p of B separates B into two components the closure of one of which is of diameter $< \epsilon$ and homeomorphic with a closed 2-cell.*[6] From 7.2 it follows immediately that exactly one component of $B - \beta$ contains points at a distance $\geqq \epsilon/2$ from p. Accordingly all simple closed curves in the complement of that component are of diameter $< \epsilon$ so that they separate B. The method of 3.3 now shows that the number of components of $B - \beta$ is two. (Take μ in the big component of $B - \beta$.) The method of 3.4 shows that an arc in the small component of $B - \beta$ spanning β separates that small component into exactly two parts. (Take β of 7.3 as $\alpha + \gamma$ of 3.4.) The method 4.1 to 4.5 is now applicable and shows that the closure of the small component is homeomorphic with a 2-cell.

7.4. We call a point of B *regular* if it has a neighborhood homeomorphic with a 2-cell. We construct first an infinite manifold covering the set C of all regular points of B.

We can find a sequence (finite if C is compact) of open 2-cells $M_1 \cdots M_n \cdots$, covering C and then a similar sequence $N_1 \cdots N_n \cdots$, $\bar{N}_n \subset M_n$, still covering C. We put the complex K_1 equal to the 2-cell N_1 with boundary. Suppose a connected complex K_{m-1} has been constructed covering $\sum_{i<m} \bar{N}_i$; we will construct a connected complex K_m covering $\sum_{i \leqq m} \bar{N}_i$.

In $M_m \cdot (C - K_{m-1})$ there is only a finite number of regions meeting N_m.

Each such region has a finite number of subregions contained in $M_m - N_m$ and having limit points on the boundaries of both M_m and N_m. In each of the smaller regions U we can find an arc α, spanning both K_{m-1} and the boundary of U, and separating in U the boundaries of M_m and N_m. We define K_m as the sum L_m of K_{m-1} and all arcs α, plus the 2-cells separated by L_m from the boundary of M_m. We exclude the trivial cases where K_{m-1} does not meet N_m or its boundary.

As no interior cell of K_{m-1} is ever subdivided and each point of C is regular this process leads to an infinite manifold covering C (if C is not compact, otherwise of course to a finite manifold).

7.5. The rest of the proof will consist in proving that $B = C$ or in other words that *every point of B is regular.*

Every point of the given simple closed curve J (IVa or Va) is limit point of $B - J$ and is a non-cut point of B.[7]

[6] In the case of Theorem V we take ϵ_F instead of ϵ where F is the compact closure of a sufficiently small neighborhood of p.

[7] See Id in 3.2. These are the only properties of J that we will use.

B contains simple closed curves in arbitrary neighborhoods of any point p of J. Take a sequence of arcs in B from a point not on J to a sequence of different points on J converging to p (see If in 3.3) and take $\eta > 0$ such that the closure of the η neighborhood of p is compact. We can find a subsequence of subarcs $p_i q_i$ of these arcs such that: the p_i are on J and converge to p; for a certain δ, $0 < \delta < \eta$, q_i is the first point on the arc starting at p_i of which the distance to p_i is $\delta/3$; the q_i converge to a point q.

If q is not on J, we can find a number m such that for $i \geqq m$, there is an arc $p\,p_i$ on J of diameter $\leqq \delta/3$ and an arc $q\,q_i$ in B of diameter $\leqq \delta/3$ not meeting J. Then for $i > m$ $p_m\,p_i + p_i\,q_i + q_i\,q + q\,q_m + q_m\,p_m$ will contain a simple closed curve contained in the η neighborhood of p.

If q is on J, we need a specification of δ left out until now. All points on J within a distance $\delta/3$ of p must lie on a subarc of J within a distance $\eta/3$ of p. There is a number m such that for $i = m$ there is an arc $q\,q_i$ of diameter $\leqq \delta/3$ not meeting p_i. Then $p_i q_i + q_i\,q +$ the subarc $q\,p_i$ of J will contain a simple closed curve contained in the η neighborhood of p.

From 7.3 and 7.5 it follows that B *contains regular points.*

7.6. Suppose B contains a point that is not regular. Then we can find a non-regular point p in B that is an end point of an arc α consisting of regular points. We only need to take an end point of an interval of regular points on an arc in B joining a regular and a non-regular point. p is on the cyclic element meeting α of any neighborhood of p. So p is on arbitrarily small simple closed curves and accordingly p is not a cut point of B.

We can find a neighborhood V of p such that every simple closed curve in V determines a 2-cell in V. We take a neighborhood U of p of diameter $< \eta\ (\epsilon/2)$ (see 7.3) and define V as the sum of U and all open 2-cells determined by simple closed curves in U. It is obvious that any simple closed curve β in V determines a 2-cell contained in V if β is contained in the sum of U and a finite number of the 2-cells added to U. But this is the case for every β. For U and the 2-cells meet β in open sets covering β and β is already covered by a finite number of such sets.

7.7. The component R of the set of all regular points in V determined by an arc of regular points in V having p as end point (α of 7.6) is homeomorphic with an open 2-cell. This follows from 7.4; compare Theorem V″.

We can find an open set W containing non-regular points, such that $R \subset W \subset \bar{R}$.

The cyclic element determined by R in V is equal to $\bar{R}\cdot V$ and each point of $(\bar{R} - R)\cdot V$ is accessible from R. For each limit point of R in V is in that cyclic element and each point q of that cyclic element is on a simple closed curve in V together with any point in R, so q is on a closed 2-cell meeting R.

We take away from V the cut points of V on $\bar{R}\cdot V$ and all parts of V separated from $\bar{R}\cdot V$ by the cut points. The rest W is an open set (of course satisfying $R \subset W \subset \bar{R}$) containing at least one non-regular point.

For $\bar{R} - R$ is a connected set containing at least one continuum of non-regular points and the cut points on $\bar{R}\cdot V$ have no limit point on V. The last

statement follows because the cut points on $\bar{R} \cdot V$ cannot be cut points of B so that the parts separated from $\bar{R} \cdot V$ must have limit points on the boundary of V. Now if the cut points had a limit point on V we could find a limit point in $V - \bar{R} \cdot V$ of a sequence of points in different components of $V - \bar{R} \cdot V$ and in that limit point B could not be locally connected.

7.8. We prove now that *the existence of such an open set W implies a contradiction.*

Any arc β in R with end points in $\bar{R} - R$ separates W. The open arc $< \beta >$ separates R into two parts. If these parts had common limit points on W, we could find an arc in W not meeting β and with end points in different parts of $R - < \beta >$. This arc could be completed to a simple closed curve in W crossing β once[8] but then one end point of β would be interior to the 2-cell in W determined by that simple closed curve. So β divides W into two parts.

If we take two non-regular points q and r in W on the same component of non-regular points in W, we can find an arc $< \beta > = < qr >$ in R such that β is part of a simple closed curve in W. We will prove that β separates B in contradiction with IVc or Vc. If both parts of $W - \beta$ had limit points on $\bar{R} - W$, we could find an arc β' crossing β once and having its end points on $\bar{R} - W$. Then $< \beta' >$ would separate W between q and r without containing non-regular points in contradiction with the assumption that q and r are on a connected set of non-regular points in W.

7.9. We draw attention to the following corollaries of Theorem V that follow from V by combinatorial considerations.

THEOREM V′. *A P-space space B satisfying the following three conditions is homeomorphic with a region on a sphere.*

V′a. *B contains at least one simple closed curve.*

V′b. *Every simple closed curve in B separates B.*

V′c. *No closed arc of a simple closed curve in B separates B.*

The complement of the region on the sphere can always be taken as a totally disconnected point set.

THEOREM V″. *A non-compact P-space satisfying the following three conditions is homeomorphic with an open 2-cell.*

V″a. *B contains at least one simple closed curve.*

V″b. *Every simple closed curve in B separates B and at least one of the components has a compact closure.*

V″c. *No closed arc of a simple closed curve in B separates B.*

8.1. We now go over to the type of characterization introduced by Whitney (14) for the closed 2-cell. The homologies and cycles occurring are meant in the Victoris sense. No specification is necessary about the modular character of the cycles. Any type will do. We do not give an introduction to the combinatorial topology used.

[8] This means that β and the simple closed curve have one point in common of which a neighborhood can be transformed into the interior of a circle in such a way that parts of β and the simple closed curve are transformed into diameters of the circle.

THEOREM VI. *A compact P-space D is homeomorphic with a 2-dimensional manifold if it contains irreducibly a 2-cycle and is separated by each simple closed curve of diameter less than* $\delta > 0$.

THEOREM VII. *A compact P-space C is homeomorphic with a 2-cell if it contains a simple closed curve J of which the fundamental 1-cycle is irreducibly homologous to zero in C and is separated by each arc spanning J.*

8.2. These theorems reduce almost immediately to preceding theorems by means of the following well known

LEMMA. *Suppose A and B are compact. Let* $C_k \to \Gamma_{k-1}$ *be a chain in* $A + B$ *and its boundary, and let U and V be arbitrary neighborhoods of the intersections of their carriers with* $A \cdot B$.

Then it is possible to find chains $C'_k \to \Gamma'_{k-1}$ *in A,* $C''_k \to \Gamma''_{k-1}$ *in B,* C^*_k *in V such that:*

$$C^*_k \to \Gamma_{k-1} - \Gamma'_{k-1} - \Gamma''_{k-1},$$

$$C_k - C'_k - C''_k - C^*_k \sim 0 \text{ in } U.$$

The upshot of this is that, after arbitrary small changes in the neighborhoods of their intersections with $A \cdot B$, the chains C_k and Γ_{k-1} can be separated into parts contained in A and B.

8.3. Under the assumptions of Theorem VI *no simple arc or point separates D.* Suppose the contrary. Then we could find two compact sets A and B such that $A + B = D$ and $A \cdot B = \alpha$, where α is an arc or a point. We can, according to the lemma, suppose that the cycle Γ_2, supposedly existing in D, is the sum of two chains C'_2 and C''_2 in A and B. If $C'_2 \to \Gamma_1$ and $C''_2 \to -\Gamma_1$, then Γ_1 is in α and so homologous to zero in α: $C^*_2 \to \Gamma_1$. Now $\Gamma_2 = (C'_2 - C^*_2) + (C''_2 + C^*_2)$ where both parts are cycles respectively in A and B. This contradicts the supposition that Γ_2 was irreducibly contained in D. It follows (see 3.2) that every component of the complement of a simple closed curve in D has every point of that simple closed curve as limit point, so that Theorem VI follows from Theorem IV (see footnote 7).

Under the assumptions of Theorem VII we can by a method analogous to that used just now prove that C is not disconnected by the sum of two arcs having no point in common and each having one end point in common with J. Now Theorem VII follows from Theorem II. However, the shortest proof of Theorem VII is much simpler than that of Theorem II, for the most complicated part of that proof can be replaced by simple consequences of the lemma. The reader is referred for this to Whitney's paper (14).

9.1. We now start the investigation of P-spaces satisfying parts of the theorems of Janiszewski. In order to emphasize the difference between the conditions in Theorems VIII and X, and Theorem XI and section 11, we do not use the word continuum. The proofs are held very short. For more details the literature should be consulted.[9]

[9] See (8) and (9) and references given there.

THEOREM VIII. *In a P-space K the following properties are equivalent:*

a. *If K is the sum of two closed and connected sets, their product is connected.*

b. *If R is a component of the complement of a closed and connected set in K, then the boundary of R is connected.*

c. *If A and B are closed, do not meet and do not separate (p from q in) K, then $A + B$ does not separate (p from q in) K.*

d. *If a and b are in closed sets A and B, $A \cdot B = 0$, then there is a closed and connected set in $K - (A + B)$ separating a and b.*

A P-space with these properties is called *unicoherent.*

9.2. *a implies b.* If R is chosen as in b, $\bar{R}$ and $K - R$ are closed and connected, $\bar{R} + (K - R) = K$, and $\bar{R} \cdot (K - R) = \bar{R} - R$. So from a it follows that $\bar{R} - R$ is connected.

b implies c. Suppose $A + B$ separates p from q. Call R the component of $K - (A + B)$ containing p and S the component of $K - \bar{R}$ containing q. Then the boundary of S is according to b a connected subset of $A + B$, separating p from q in K. If A and B are closed and do not meet, that boundary is contained in A or B, so that A or B separates p from q in K.

c implies d. Suppose A and B closed and separated, a in A, b in B. $K - (A + B)$ contains a closed set separating a and b, and that contains an irreducible closed set C separating a and b. C must be connected according to c.

d implies a. Suppose $K = K_1 + K_2$, K_1 and K_2 closed and connected, $K_1 \cdot K_2 = A + B$, A and B separated. We can find a closed and connected set C in $K - (A + B)$ separating a point in A from a point in B. C would have to meet K_1 and K_2 in separated sets of which it is the sum, and that is impossible.

In c the bracketed words can be left out. Suppose A and B are closed, do not meet and do not separate p from q. Form A' by adding to A all components of $K - A$ not containing $p + q$, and form B' by leaving out parts of B contained in those components. Then form A'' and B'' by the same process with A' and B' instead of B and A. Now A'' and B'' do not separate K so that according to the weaker form of c $A'' + B''$ containing $A + B$ does not separate K so that $A + B$ does not separate p from q in K. And this last statement implies the stronger form of c.

9.3. THEOREM XI. *In a P-space K the following four properties are equivalent:*

a′. *If K is the sum of two closed and connected sets of which the product is compact, then their product is connected.*

b′. *If R is a component of the complement of a closed and connected set in K, and the boundary of K is compact, then it is connected.*

c′. *If A and B are compact, neither separates p from q in K, and $A \cdot B = 0$, then $A + B$ does not separate p from q in K.*

d′. *If a and b are in compact sets A and B, $A \cdot B = 0$, then there is a compact connected set in $K - (A + B)$ separating a and b.*

We leave out the proofs because they are very similar to those of Theorem VIII. Obviously if a P-space is unicoherent it has the properties of Theorem XI. In a compact Peano space all eight properties are equivalent.

10.1. THEOREM X. *In a P-space K the following four properties are equivalent:*

a. *If the closed and connected set C does not separate K, then C is unicoherent.*

b. *If a closed set C separates K and C is the sum of two closed sets C_1 and C_2 that do not separate K, then $C_1 \cdot C_2$ is not connected.*

c. *If a closed set C separates p from q and is the sum of two closed sets C_1 and C_2 that do not separate p from q, then $C_1 \cdot C_2$ is not connected.*

d. *Each non-degenerate cyclic element of K is homeomorphic with a 2-sphere.*

10.2. Of course, *c implies b*. We prove that *b implies c*. Take sets C, C_1, C_2 satisfying the conditions of c. Form C_1' by adding to C_1 all components of $K - C_1$ not containing $p + q$, and form C_2' by subtracting from C_2 its intersection with the sum of those components. C_1' does not separate K, $C_1' + C_2' \supset C_1 + C_2$ and $C_1' \cdot C_2' = C_1 \cdot C_2$. Now form C_2'' by adding to C_2' all components of $K - C_2'$ not containing $p + q$ and form C_2'' by subtracting from C_1' its intersection with the sum of those components of $K - C_2'$ that do not contain $p + q$ and do not belong entirely to C_1'. Then $C_1'' + C_2'' \supset C_1' + C_2'$ so that it separates p from q. Neither C_1'' nor C_2'' separates K (for C_1'' this follows because any point taken away from C_1' was in a connected set with points not belonging to C_1'). Finally $C_1'' \cdot C_2'' - C_1' \cdot C_2'$ consists of regions in K of which the boundary belonged to both C_1' and C_2', so that the number of components of $C_1'' \cdot C_2''$ is not more than the number of components of $C_1' \cdot C_2' = C_1 \cdot C_2$. We can apply b on the new C_1'' and C_2'' finding c for the sets C_1 and C_2.

10.3. *a implies c.* Suppose the closed sets C_1 and C_2 do not separate p from q and $C_1 \cdot C_2$ is connected; we prove that $C_1 + C_2$ does not separate p from q. We can find arcs B_1 and B_2 joining p and q in $K - C_1$ and $K - C_2$. Call B the complement of the component of $K - (B_1 + B_2)$ containing $C_1 \cdot C_2$. B is unicoherent according to a. $B \cdot C_1$ and $B \cdot C_2$ are closed and separated and neither separates p from q in B. So $B \cdot (C_1 + C_2)$ does not separate p from q in B and then $C_1 + C_2$ cannot separate p from q in K.

10.4. *c implies a′: If the compact and connected set C does not separate K, then C is unicoherent.* a′ is a weaker form of a, identical with a if K is compact.

We suppose that the compact and connected set C is the sum of two compact and connected sets C_1 and C_2 and that $C_1 \cdot C_2$ is the sum of two separated sets A and B and we prove that C separates K. K is unicoherent (see Theorem VIIIc). Accordingly points a in A and b in B can be separated by a compact and connected set M not meeting $A + B$. M intersects C_1 and C_2 in separated sets M_1 and M_2. Take two neighborhoods N_1 and N_2 of M_1 and M_2 in M, $N_1 \cdot N_2 = 0$. If C does not separate K we can find a compact connected set N in $K - C$ containing $M - (N_1 + N_2)$. Now neither $N_1 + N$ nor $N_2 + N$ separates K between a and b, but their sum does because it contains M and their product N is connected. This contradicts c.

10.5. *c implies d.* If K satisfies c, then each non-degenerate cyclic element L of K satisfies c. For if a set does or does not separate p from q in L, it does the same in K. From c it follows that no closed arc separates L because a separating arc would have to contain an irreducible connected separating set

and this cannot be an arc according to c and not a point because L does not have cut points. L satisfies a′ because it satisfies c so that each simple closed curve separates L. It follows first that L is homeomorphic with a region on the sphere, and then, because we can now prove that no open arc separates, that L is homeomorphic with a sphere.

10.6. *A 2-sphere L has properties a and b.* For a compact space a and b follow from c by 10.4 and 10.2. So we prove c. Suppose the two closed sets C_1 and C_2 do not separate two points p and q and have a connected intersection. We prove that $C_1 + C_2$ does not separate p and q. Join p to q by two polygonal arcs B_1 and B_2 in $L - C_1$ and $L - C_2$. The complement B of the component of $L - (B_1 + B_2)$ containing $C_1 \cdot C_2$ is a 2-cell and we only have to prove that this 2-cell satisfies VIIIc. This follows as the reader can verify immediately by proving VIIId.

d implies a. We must prove: if all non-degenerate cyclic elements of a P-space K have property a, then the space itself has property a. A connected and closed set C not separating K does not separate any cyclic element L of K, and its intersection with L is closed and connected. We still have to prove: if a closed and connected set C in K has unicoherent intersections with each cyclic element of K, then C itself is unicoherent. Suppose C is the sum of two closed and connected sets C_1 and C_2. That $C_1 \cdot C_2$ is connected follows because its intersection with each cyclic element of K is connected and because if two points belong to $C_1 \cdot C_2$ all cut points of K between these two points belong to both C_1 and C_2. That means they belong to $C_1 \cdot C_2$.

11.1. The following conditions on a P-space K similar to those in Theorem X do not have quite such a simple relationship.

a′. If the compact and connected set C does not separate K, then C is unicoherent.

b′. If a compact set C separates K and C is the sum of two closed sets C_1 and C_2 that do not separate K, then $C_1 \cdot C_2$ is not connected.

c′. If a closed set C separates p from q and is the sum of two closed sets C_1 and C_2 that do not separate p from q, then $C_1 \cdot C_2$ is not connected.

d′. Each non-degenerate cyclic element of K is homeomorphic with a region on a 2-sphere.

c′ and d′ are equivalent and imply a′ and b′. The proofs are similar to those of sections 10.4 to 10.6 and are left to the reader. Theorem IX is used instead of Theorem VIII. *a′ does not imply either one of the others* as shown by a closed 2-cell from which one boundary point has been removed. It remains an open question whether or not b′ implies any of the other properties.

11.2. *a′ and b′ together are equivalent to c′ or d′.* We give a short sketch of the proof that a′ + b′ implies d′.

Suppose that K satisfies a′ and b′. A cyclic element L of K cannot have a local cut point p, for otherwise we could find in a sufficiently small neighborhood of p two compact sets, not separating K, not having a common point, but of

which the sum separates K. If we take away from L all cut points of K on L that separate non-compact sets from L, then the rest is a cyclic p-space M satisfying a′ and b′. M is open in K because the cut points taken away cannot have a limit point. M is connected and cyclically connected because otherwise we could find a local cut point of L. M satisfies a′: if a compact and connected set C does not separate M, the cut points of K in C will separate from M a compact point set D. $C + D$ does not separate K and C is unicoherent if $C + D$ is. M satisfies b′: if a compact set C separates M it separates K. For if it did not we could find a common limit point in K of the separated parts of $M - C$; that point would be a local cut point of L and that is impossible.

From b′ it follows that no arc separates M and from a′ that every simple closed curve separates M, so that M is homeomorphic with a region on a sphere. The complement of that region can be supposed totally disconnected. To any sequence in M converging to a point of $L - M$ there must correspond a convergent sequence on the sphere for otherwise that point of $L - M$ would be a local cut point of M. So L too is homeomorphic with a region on the sphere.

APPENDIX

1. We will prove that the generalized manifolds of E. Čech and S. Lefschetz do not include other spaces than manifolds in the 2-dimensional case.[10] A connected 2-dimensional generalized manifold M has the following properties:

a. M is a Peano space (C, Axioms A and G_0; L, p. 487, III).

b. M is an open 2-circuit (C, 17 and 17.5; L, p. 487, I).

c. M is locally 1-connected (C, Axiom G_1; L, p. 488, 5th line; the statement there that the condition is included in IV is not correct).

d. For each point x of M there is a 2-chain p_x such that for any 2-chain r_2 in M of which the boundary does not meet x there is a rational number m such that $r_2 - mp_x$ is homologous to a chain not meeting x (C, Axiom E; L, p. 487, II).

We will prove that these conditions imply that M is an open 2-dimensional manifold.

2. As in 8.3 we can derive from a and b that M is not separated by a finite number of points or a closed arc. So M contains a simple closed curve α and the closure of the complement of α contains α. The reasoning of 7.5 now proves that M contains arbitrarily small simple closed curves. Remembering Theorem V and condition c we see that M must be a manifold as a result of the following lemma:

[10] E. Čech, *Théorie générale des variétés et de leurs théorèmes de dualité*, Annals of Math., vol. 34 (1933), pp. 621–730—"C"; S. Lefschetz, *On generalized manifolds*, Am. Journal of Math., vol. 55 (1933), pp. 469–504—"L". We will give our proof on the basis of L. C uses a different definition of homology and does not explicitly assume that the spaces considered are separable metric. We will not prove in detail that passage from the one treatment to the other is possible, but cite the paragraphs of C where properties corresponding to the ones used can be found. It is to be remarked that in both papers the field of rational numbers is used as field of coefficients.

If the covering 1-*cycle* p_1 *of an s.c.c.* α *in* M *is homologous to zero in* M, *then* α *separates* M. Suppose the 2-chain p_2 has p_1 as boundary and q_2 is the covering 2-cycle of M. In each point x of $M - \alpha$, both p_2 and q_2 determine a rational multiple of the chain p_x postulated in d and the second multiple is different from zero. Otherwise q_2 would be homologous to a cycle not meeting x. So in each point x of $M - \alpha$ we can determine a rational quotient m_x of p_2 and q_2. As we can replace p_2 by $p_2 - nq_2$, where n is any rational number, we can replace m_x by $m_x - n$ for all x of $M - \alpha$. So we can suppose that for a certain point x the number m_x is equal to zero. From d we can then conclude that p_1 is still homologous to zero in some proper subset of M, for the chain p_2 determines then the zero multiple of p_x in the point x. Furthermore we see that if $m_x = 0$, then $m_y = 0$ for all points of some neighborhood of x so that the set of points x in which m_x has a certain constant value is open in M.

We take a subset N of M in which p_1 is irreducibly homologous to zero. In order to prove that α separates M, it is sufficient to show that $N_1 = M - N$ cannot have any limit points on $N_2 = N - \alpha$. For then N_1 and N_2, neither of which can be vacuous, are separated by α.

Suppose x is a limit point of N_1 contained in N_2. Then the number m_x for the chain $p_2 \to p_1$ in N is equal to zero because any neighborhood of x contains points of N_1 in which the number is obviously zero. So p_2 is homologous to a chain r_2 not meeting x. As M is 2-dimensional it cannot contain a 3-chain essentially different from zero, so that a 3-chain r_3 exists in the carrier of $r_2 - p_2$ with boundary $r_2 - p_2$. Now we add to p_2 the boundary of the part of r_3 in a neighborhood of x that does not meet r_2. Then the resulting 2-chain is contained in a proper subset of N and has p_1 as boundary so that p_1 was not *irreducibly* homologous to zero in N. This contradiction proves that N_1 and N_2 are parts of N separated by α.

LITERATURE

1. R. L. Moore, *On the foundations of plane analysis situs*, Tr. Am. Math. Soc., vol. 20 (1919), pp. 169–178.
2. R. L. Moore, *Concerning a set of postulates for plane analysis situs*, Tr. Am. Math. Soc., vol. 17 (1916), pp. 131–164.
3. R. L. Moore and J. R. Kline, Bull. Am. Math. Soc., vol. 28 (1922), p. 380, Abstr. 8.
4. T. Radó, *Über den Begriff der Riemannschen Fläche*, Acta litt. ac scient., vol. 2 (1925), pp. 101–121.
5. I. Gawehn, *Über unberandete 2-dimensionale Mannigfaltigkeiten*, Math. Ann., vol. 98 (1928), pp. 321–354.
6. C. Kuratowski, *Un système d'axiomes pour la topologie de la surface de la sphère*, Congress Bologna, 1928, vol. IV, p. 239.
7. D. W. Woodard, *On two-dimensional analysis situs with special reference to the Jordan curve theorem*, F. M., vol. 13 (1929), pp. 121–145.
8. C. Kuratowski, *Une caractérization de la surface de la sphère*, F. M., vol. 13 (1929), pp. 307–318.
9. L. Zippin, *A study of continuous curves and their relation to the Janiszewski-Mullikin theorem*, Tr. Am. Math. Soc., vol. 31 (1929), pp. 744–770.
10. R. L. Wilder, Bull. Am. Math. Soc., vol. 35 (1929), p. 194; Abstr. 8.

11. R. L. Wilder, *A converse of the Jordan-Brouwer separation theorem in three dimensions*, Tr. Am. Math. Soc., vol. 32 (1930), pp. 632–651.
12. L. Zippin, *On continuous curves and the Jordan curve theorem*, Am. J. of Math., vol. 52 (1930), pp. 331–350.
13. J. H. Roberts, *A point-set characterization of closed 2-dimensional manifolds*, F. M., vol. 18 (1932), pp. 37–46.
14. H. Whitney, *A characterization of the closed 2-cell*, Tr. Am. Math. Soc., vol. 35 (1933), pp. 261–273.
15. L. Zippin, *A characterization of the closed 2-cell*, Am. J. of Math., vol. 55 (1933), pp. 207–217.

The Johns Hopkins University.

HEAT CONDUCTION IN A SEMI-INFINITE SOLID OF TWO DIFFERENT MATERIALS

By Arnold N. Lowan

Consider the unidimensional flow of heat in a semi-infinite solid extending from $x = -a$ to $x = \infty$. Assume that the section extending from $x = -a$ to $x = 0$ is initially at some given temperature $f_1(x)$ and has thermal diffusivity k_1, while the remaining solid is initially at temperature $f_2(x)$ and has thermal diffusivity k_2. If the boundary $x = -a$ is permanently kept at the temperature T_0, what is the subsequent thermal history of the solid?

A solution of this problem with the aid of contour integrals is given by Carslaw[1] in the special case $f_1(x) = f_2(x) = 0$. The case of nonvanishing initial temperatures could be dealt with by the general method of Green's functions, but as far as the author is aware, the complete treatment of the problem by this method is not to be found in the literature. The method to be employed here is essentially that presented in a number of earlier papers,[2] to which the reader is referred for the derivation of some of the results to be subsequently used.

For the sake of simplicity, we shall assume that $T_0 = 0$. It is clear that this implies no loss of generality. The mathematical formulation of the problem is as follows:

$$\frac{\partial}{\partial t} T_1(x, t) - k_1 \frac{\partial^2}{\partial x^2} T_1(x, t) = 0 \qquad -a < x < 0\,, \tag{1}$$

$$\lim_{t \to 0} T_1(x, t) = f_1(x)\,, \tag{2}$$

$$T_1(-a, t) = 0\,, \tag{3}$$

$$\frac{\partial}{\partial t} T_2(x, t) - k_2 \frac{\partial^2}{\partial x^2} T_2(x, t) = 0 \qquad 0 < x < \infty\,, \tag{4}$$

$$\lim_{t \to 0} T_2(x, t) = f_2(x)\,, \tag{5}$$

$$K_1 \frac{\partial}{\partial x} T_1(x, t) = K_2 \frac{\partial}{\partial x} T_2(x, t) \qquad x = 0\,, \tag{6}$$

$$T_1(0, t) = T_2(0, t)\,. \tag{7}$$

Received by the Editors of the American Journal of Mathematics, September 4, 1934, accepted by them, and later transferred to this journal.

[1] H. S. Carslaw, *Mathematical Theory of the Conduction of Heat in Solids*, 1926.

[2] A. N. Lowan, *Cooling of a radioactive sphere*, Physical Review, vol. 44 (1933), pp. 769–775, cited as L1.

Heat conduction in a semi-infinite solid, Amer. Jour. Math., vol. 56 (1934), pp. 396–400, cited as L2.

On the problem of the heat recuperator, Phil. Mag., vol. 17 (1934), pp. 914–933, cited as L3.

If we designate by $\varphi(t)$ the common value of the two members of (6), where we assume provisionally that $\varphi(t)$ is a known function, then the system A, consisting of (1), (2), (3) and the boundary condition

$$K_1 \frac{\partial}{\partial x} T_1(x, t) = \varphi(t) \qquad x = 0 \tag{6'}$$

may be solved by the method presented in L1. Similarly, the system B, consisting of (4), (5) and the boundary condition

$$K_2 \frac{\partial}{\partial x} T_2(x, t) = \varphi(t) \qquad x = 0 \tag{6''}$$

may be solved with the aid of the result obtained in L2. If the solutions of the systems A and B are substituted in (7), we obtain a Volterra integral equation, which may be solved by the method employed in L3.

We proceed to solve the system A. Let us make the substitution

$$T_1(x, t) = u(x, t) + \frac{x + a}{K_1} \varphi(t). \tag{8}$$

It is then readily seen that the function $u(x, t)$ must satisfy the system

$$\frac{\partial}{\partial t} u(x, t) - k_1 \frac{\partial^2}{\partial x^2} u(x, t) = -\frac{x + a}{K_1} \varphi'(t) = \phi_1(x, t), \qquad \text{(say)} \tag{9}$$

$$\lim_{t \to 0} u(x, t) = f(x) - \frac{x + a}{K_1} \varphi(0) = F_1(x), \qquad \text{(say)} \tag{10}$$

$$u(-a, 0) = 0, \tag{11}$$

$$\frac{\partial}{\partial x} u(x, t) = 0 \qquad x = 0. \tag{12}$$

Reference to L1 and L2 shows clearly that the system C, consisting of equations (9) to (12), may be said to be the mathematical formulation of the problem of the flow of heat in a radioactive slab, with the radioactivity function $\phi_1(x, t)$. Making use of the results of L1 and L2, the solution of the system C may be written down at once in the form

$$\begin{aligned} u(x, t) = {} & \sum_{n=0}^{\infty} y_n(x)\, e^{-k_1 \lambda_n^2 t} \int_{-a}^{0} F_1(\xi)\, y_n(\xi)\, d\xi \\ & + \sum_{n=0}^{\infty} y_n(x)\, e^{-k_1 \lambda_n^2 t} \int_{-a}^{0} y_n(\xi) \left\{ \int_0^t \phi_1(\xi, \eta)\, e^{k_1 \lambda_n^2 \eta}\, d\eta \right\} d\xi, \end{aligned} \tag{13}$$

where the summation extends over the characteristic values and the corresponding normalized characteristic functions of the homogeneous differential equation

$$y'' + \lambda^2 y = 0 \tag{14}$$

in conjunction with the boundary conditions (11) and (12).

Clearly

(15) $$\lambda_n = (2n + 1)\,\pi/2a\,, \qquad y_n(x) = \sqrt{2/a}\,\cos \lambda_n x\,.$$

In view of (13) and of the significance of the functions $F_1(x)$ and $\phi_1(x, t)$ from (9) and (10), equation (8) becomes

(16) $$\begin{aligned} T_1(x, t) = \frac{(x+a)}{K_1}\,\varphi(t) &+ \sum_{n=0}^{\infty} y_n(x)\, e^{-k_1\lambda_n^2 t} \int_{-a}^{0} f_1(\xi)\, y_n(\xi)\, d\xi \\ &- \frac{\varphi(0)}{K_1} \sum_{n=0}^{\infty} y_n(x)\, e^{-k_1\lambda_n^2 t} \int_{-a}^{0} (\xi + a)\, y_n(\xi)\, d\xi \\ &- \frac{1}{K_1} \sum_{n=0}^{\infty} y_n(x)\, e^{-k_1\lambda_n^2 t} \int_{-a}^{0} (\xi + a)\, y_n(\xi) \left\{ \int_0^t \varphi'(\eta)\, e^{k_1\lambda_n^2 \eta}\, d\eta \right\} d\xi\,. \end{aligned}$$

If the last integral is integrated by parts and we make use of the identity

$$x + a \equiv \sum_{n=0}^{\infty} y_n(x) \int_{-a}^{0} (\xi + a)\, y_n(\xi)\, d\xi\,,$$

(16) ultimately becomes

(17) $$\begin{aligned} T_1(x, t) = &\sum_{n=0}^{\infty} y_n(x)\, e^{-k_1\lambda_n^2 t} \int_{-a}^{0} f_1(\xi)\, y_n(\xi)\, d\xi \\ &+ \frac{k_1}{K_1} \sum_{n=0}^{\infty} \lambda_n^2\, y_n(x)\, e^{-k_1\lambda_n^2 t} \int_{-a}^{0} (\xi + a)\, y_n(\xi)\, d\xi \int_0^t \varphi(\eta)\, e^{k_1\lambda_n^2 \eta}\, d\eta\,. \end{aligned}$$

This is the complete solution of the system A. For $x = 0$ (17) yields

(18) $$T_1(0, t) = A_1(t) + \int_0^t \varphi(\eta)\, M_1\,(t - \eta)\, d\eta\,,$$

where

(19) $$A_1(t) = \sum_{n=0}^{\infty} y_n(0)\, e^{-k_1\lambda_n^2 t} \int_{-a}^{0} f_1(\xi)\, y_n(\xi)\, d\xi\,,$$

(20) $$M_1(t) = \frac{k_1}{K_1} \sum_{n=0}^{\infty} \lambda_n^2\, y_n(0)\, e^{-k_1\lambda_n^2 t} \int_{-a}^{0} (\xi + a)\, y_n(\xi)\, d\xi\,.$$

We now proceed to solve the system B. If we make the substitution

(21) $$T_2\,(x, t) = v(x, t) + \frac{x}{K_2}\,\varphi(t)\,,$$

it is clear that the function $v(x, t)$ must satisfy the system of equations

$$(22)\qquad \frac{\partial}{\partial t} v(x, t) - k_2 \frac{\partial^2}{\partial x^2} v(x, t) = -\frac{x}{K_2} \varphi'(t) = \phi_2(x, t), \qquad \text{(say)}$$

$$(23)\qquad \lim_{t \to 0} v(x, t) = f_2(x) - \frac{x}{K_2} \varphi(0) = F_2(x), \qquad \text{(say)}$$

$$(24)\qquad \frac{\partial}{\partial x} v(x, t) = 0 \qquad x = 0.$$

As in the previous case, the system D of the last three equations may be said to be the mathematical formulation of the problem of the flow of heat in a semi-infinite radioactive solid.

This problem has been treated in L2 for the most general case where radiation takes place at the boundary plane. The boundary condition (24) may be obtained from that in L2 by putting the coefficient of heat transfer $h = 0$. Accordingly, the solution of D is obtained from that of L2 (by putting in the latter $h = 0$) in the form

$$\begin{aligned} v(x, t) = {} & \frac{1}{2\sqrt{\pi k_2 t}} \int_0^\infty \int_0^\infty \left\{ \exp\left[-\frac{(x + \rho + \xi)^2}{4k_2 t} \right] \right. \\ & \left. - \exp\left[-\frac{(x + \rho - \xi)^2}{4k_2 t} \right] \right\} \frac{\partial F_2}{\partial \xi}\, d\xi\, d\rho \\ (25)\qquad & + \frac{1}{2\sqrt{\pi k_2}} \int_0^\infty d\xi \int_0^\infty d\rho \int_0^\infty (t - \eta)^{-\frac{1}{2}} \frac{\partial \phi_2}{\partial \xi} \left\{ \exp\left[-\frac{(x + \rho + \xi)^2}{4k_2(t - \eta)} \right] \right. \\ & \left. - \exp\left[-\frac{(x + \rho - \xi)^2}{4k_2(t - \eta)} \right] \right\} d\eta \\ = {} & v_1(x, t) + v_2(x, t), \quad \text{say}. \end{aligned}$$

Consider the double integral

$$I_1 = \int_0^\infty d\rho \int_0^\infty \frac{\partial F_2}{\partial \xi} \exp\left[-\frac{(x + \rho + \xi)^2}{4k_2 t} \right] d\xi.$$

If the second integral is integrated by parts, we ultimately get

$$(26)\quad I_1 = -f_2(0) \int_0^\infty \exp\left[-\frac{(x + \rho)^2}{4k_2 t} \right] d\rho + \int_0^\infty F_2(\xi) \exp\left[-\frac{(x + \xi)^2}{4k_2 t} \right] d\xi.$$

Similarly we get

$$(27)\quad J_1 = -f_2(0) \int_0^\infty \exp\left[-\frac{(x + \rho)^2}{4k_2 t} \right] d\rho - \int_0^\infty F_2(\xi) \exp\left[-\frac{(x - \xi)^2}{4k_2 t} \right] d\xi,$$

where J_1 is the double integral obtained from I_1 by replacing in the exponential ξ by $-\xi$. In view of (26) and (27) and of the significance of $F_2(x)$ we get

$$
(28)\quad \begin{aligned} v_1(x, t) &= \frac{1}{2\sqrt{\pi k_2 t}} \int_0^\infty f(\xi) \left\{ \exp\left[-\frac{(x+\xi)^2}{4k_2 t} \right] + \exp\left[-\frac{(x-\xi)^2}{4k_2 t} \right] \right\} d\xi \\ &- \frac{\varphi(0)}{2\,K_2 \sqrt{\pi k_2 t}} \int_0^\infty \xi \left\{ \exp\left[-\frac{(x+\xi)^2}{4k_2 t} \right] + \exp\left[-\frac{(x-\xi)^2}{4k_2 t} \right] \right\} d\xi\,. \end{aligned}
$$

By analogy with the above manipulations, which have yielded (28), it is easily seen that the second term in (25) may be written

$$
(29)\quad \begin{aligned} v_2(x, t) = -\frac{1}{2K_2\sqrt{\pi k_2}} \int_0^t \varphi'(\eta) \left\{ \int_0^\infty \frac{\xi d\xi}{\sqrt{t-\eta}} \left[\exp\left(-\frac{(x+\xi)^2}{4k_2\,(t-\eta)} \right) \right.\right. \\ \left.\left. + \exp\left(-\frac{(x-\xi)^2}{4k_2\,(t-\eta)} \right) \right] \right\} d\eta\,. \end{aligned}
$$

If the first integral is integrated by parts, (29) becomes

$$
(30)\quad \begin{aligned} v_2(x, t) = &-\frac{\varphi(t)}{K_2} \lim_{t-\eta\to 0} \int_0^\infty \frac{\xi\, d\xi}{2\sqrt{\pi k_2(t-\eta)}} \left\{ \exp\left(-\frac{(x+\xi)^2}{4k_2(t-\eta)} \right) \right. \\ &\left. - \exp\left(-\frac{(x-\xi)^2}{4k_2(t-\eta)} \right) \right\} \\ &+ \frac{\varphi(0)}{2K_2\sqrt{\pi k_2 t}} \int_0^\infty \xi \left\{ \exp\left(-\frac{(x+\xi)^2}{4k_2 t} \right) + \exp\left(-\frac{(x-\xi)^2}{4k_2 t} \right) \right\} d\xi \\ &+ \frac{1}{2K_2\sqrt{\pi k_2}} \int_0^t \varphi(\eta) \frac{\partial}{\partial\eta} \left\{ \int_0^\infty \frac{\xi\, d\xi}{\sqrt{t-\eta}} \left[\exp\left(-\frac{(x+\xi)^2}{4k_2(t-\eta)} \right) \right.\right. \\ &\left.\left. + \exp\left(-\frac{(x-\xi)^2}{4k_2(t-\eta)} \right) \right] \right\} d\eta\,. \end{aligned}
$$

In view of the well-known identity[3]

$$
(31)\quad f(x) \equiv \lim_{t\to 0} \frac{1}{\sqrt{2\pi k_2 t}} \int_0^\infty f(\xi) \left\{ \exp\left(-\frac{(x+\xi)^2}{4k_2 t} \right) + \exp\left(-\frac{(x-\xi)^2}{4k_2 t} \right) \right\} d\xi\,,
$$

the first term in (30) is $= -x\varphi(t)/K_2$. Furthermore, the second term of (30) is identical with the second term of (28), except for its sign. We, therefore, get from (21), (28) and (30)

$$
(32)\quad \begin{aligned} T_2(x, t) = &\frac{1}{2\sqrt{\pi k_2 t}} \int_0^\infty f_2(\xi) \left\{ \exp\left(-\frac{(x+\xi)^2}{4k_2 t} \right) + \exp\left(-\frac{(x-\xi)^2}{4k_2 t} \right) \right\} d\xi \\ &+ \frac{1}{2K_2\sqrt{\pi k_2}} \int_0^t \varphi(\eta) \frac{\partial}{\partial\eta} \left\{ \int_0^\infty \frac{\xi d\xi}{\sqrt{t-\eta}} \left[\exp\left(-\frac{(x+\xi)^2}{4k_2(t-\eta)} \right) \right.\right. \\ &\left.\left. + \exp\left(-\frac{(x-\xi)^2}{4k_2(t-\eta)} \right) \right] \right\} d\eta\,. \end{aligned}
$$

[3] Carslaw, loc. cit. (footnote 1), art. 18.

This is the complete solution of the system B. For $x = 0$, (32) yields

$$T_2(0, t) = A_2(t) + \int_0^t \varphi(\eta) \frac{\partial}{\partial \eta} M_2(t - \eta)\, d\eta , \tag{33}$$

where we have put

$$A_2(t) = \frac{1}{\sqrt{\pi k_2 t}} \int_0^\infty f_2(\xi) \exp\left(-\frac{\xi^2}{4k_2 t}\right) d\xi , \tag{34}$$

$$M_2(t) = \frac{1}{K_2\sqrt{\pi k_2}} \int_0^\infty \frac{\xi}{\sqrt{t}} \exp\left(-\frac{\xi^2}{4k_2 t}\right) d\xi . \tag{35}$$

Substituting (17) and (32) in (7) we get

$$A_1(t) - A_2(t) + \int_0^t \varphi(\eta) \left\{ M_1(t - \eta) - \frac{\partial}{\partial \eta} M_2(t - \eta) \right\} d\eta = 0 . \tag{36}$$

This is a Volterra integral equation for the unknown function $\varphi(t)$.

As previously stated, a solution of (36) may be obtained by means of the method employed in L3. Let us operate on (6) by the Laplace operator defined by

$$L\,\{\varphi(t)\} = \int_0^\infty e^{-pt} \varphi(t)\, dt = y(p) , \qquad \text{say} . \tag{37}$$

With this definition of the operator L, let

$$\begin{aligned} L\,\{A_1(t)\} &= A_1^*(p) , & L\,\{A_2(t)\} &= A_2^*(p) , \\ L\,\{M_1(t)\} &= M_1^*(p) , & L\,\{M_2(t)\} &= M_2^*(p) . \end{aligned} \tag{38}$$

Also let

$$N_2(t - \eta) = \frac{\partial}{\partial \eta} M_2(t - \eta) . \tag{39}$$

In view of the developments in L3 (pp. 918, 919) we have

$$L\left\{\int_0^t \varphi(\eta)\, M_1(t - \eta)\, d\eta\right\} = L\{\varphi(t)\} \cdot L\{M_1(t)\} = y(p)\, M_1^*(p) . \tag{40}$$

Similarly

$$\begin{aligned} L\left\{\int_0^t \varphi(\eta)\, N_2(t - \eta)\, d\eta\right\} &= L\{\varphi(t)\} \cdot \int_0^\infty e^{-p\tau} N_2(\tau)\, d\tau \\ &= y(p) \int_0^\infty e^{-p\tau} N_2(\tau)\, d\tau . \end{aligned} \tag{41}$$

From (39) it is clear that

$$N_2(\tau) = -\frac{\partial}{\partial \tau} M_2(\tau) , \tag{39'}$$

and, therefore, with the aid of identity (8) in L1

$$-L\left\{\frac{\partial M_2}{\partial \tau}\right\} = -p\,L\{M_2(\tau)\} + M_2(0) = -p\,M_2^*(p),$$

since comparison between (35) and (31) shows clearly that $M_2(0) = 0$. Equation (41), therefore, becomes

$$\text{(41')} \qquad L\left\{\int_0^t \varphi(\eta)\, N_2(t-\eta)\, d\eta\right\} = -p\,y(p)\,M_2^*(p).$$

The result of operating on (36) by the Laplace operator is, consequently,

$$A_1^*(p) - A_2^*(p) = y(p)\,\{M_1^*(p) + pM_2^*(p)\},$$

whence

$$\text{(42)} \qquad y(p) = \frac{A_1^*(p) - A_2^*(p)}{M_1^*(p) + p\,M_2^*(p)}.$$

We now proceed to the evaluation of the "starred" terms in (42). From (19) and (20) we get at once, since $L\{e^{-at}\} = 1/(p+a)$,

$$\text{(43)} \qquad A_1^*(p) = \sum_{n=0}^{\infty} \frac{E_n}{k_1\lambda_n^2 + p},$$

$$\text{(44)} \qquad M_1^*(p) = \sum_{n=0}^{\infty} \frac{F_n}{k_1\lambda_n^2 + p},$$

where

$$\text{(45)} \qquad E_n = y_n(0)\int_{-a}^{0} f(\xi)\, y_n(\xi)\, d\xi,$$

$$\text{(46)} \qquad F_n = \frac{k_1}{K_1}\lambda_n^2\, y_n(0)\int_{-a}^{0} (\xi + a)\, y_n(\xi)\, d\xi = \frac{2k_1}{aK_1}.$$

From (34) we get

$$\text{(47)} \qquad A_2^*(p) = \frac{1}{\sqrt{k_2}}\int_0^{\infty} f_2(\xi)\left\{\int_0^{\infty} e^{-pt}\frac{1}{\sqrt{\pi t}}\exp\left(-\frac{\xi^2}{4k_2 t}\right) dt\right\} d\xi.$$

Consider the identity[4]

$$\text{(48)} \qquad \int_0^{\infty} e^{-pt}\frac{e^{-\frac{\lambda}{t}}}{\sqrt{\pi t}}\, dt = \frac{e^{-2\sqrt{\lambda p}}}{\sqrt{p}}.$$

The value of the second integral in (47) is obtained at once from (48) by replacing λ by $\xi^2/(4k_2)$. Thus, finally

[4] J. R. Carson, *Electric Circuit Theory*, 1926, p. 39, formula (9).

(49)
$$A_2^*(p) = \frac{1}{\sqrt{k_2 p}} \int_0^\infty f_2(\xi) \exp\left(-\xi\sqrt{\frac{p}{k_2}}\right) d\xi .$$

In entirely similar fashion

(50)
$$M_2^*(p) = \frac{1}{K_2\sqrt{k_2 p}} \int_0^\infty \xi \exp\left(-\xi\sqrt{\frac{p}{k_2}}\right) d\xi .$$

In view of (43), (44), (49) and (50), equation (42) becomes

(51)
$$y(p) = \frac{\displaystyle\sum_{n=0}^{\infty} \frac{E_n}{k_1\lambda_n^2 + p} - \frac{1}{\sqrt{k_2 p}} \int_0^\infty f_2(\xi) \exp\left(-\xi\sqrt{\frac{p}{k_2}}\right) d\xi}{\displaystyle\sum_{n=0}^{\infty} \frac{F_n}{k_1\lambda_n^2 + p} + \frac{p}{K_2\sqrt{k_2 p}} \int_0^\infty \xi \exp\left(-\xi\sqrt{\frac{p}{k_2}}\right) d\xi} .$$

If we make the substitution $\xi = \sqrt{k_2/p}\;\eta$, the last equation yields

(52)
$$py(p) = \frac{\displaystyle\sum_{n=0}^{\infty} \frac{p^{3/2}E_n}{k_1\lambda_n^2 + p} - p^{1/2} \int_0^\infty f_2\left(\sqrt{\frac{k_2}{p}}\,\eta\right) e^{-\eta}\, d\eta}{\displaystyle p^{1/2}\sum_{n=0}^{\infty} \frac{F_n}{k_1\lambda_n^2 + p} + \frac{\sqrt{k_2}}{K_2}} = \frac{Y(p)}{Z(p)},$$

say.

Consider the expression

(53)
$$s(p) = \sum_{n=0}^{\infty} \frac{p^{1/2}F_n}{k_1\lambda_n^2 + p} = \frac{2\,k_1}{a\,K_1} \sum_{n=0}^{\infty} \frac{p^{1/2}}{[(2n+1)\pi/2a]^2 + p} .$$

If we make the substitution $p = k_1 q^2/a^2$, we get

(53′)
$$s(p) = \frac{\sqrt{k_1}}{K_1} \sum_{n=0}^{\infty} \frac{2q}{[(2n+1)\pi/2]^2 + q^2} = \frac{\sqrt{k_1}}{K_1} \tanh q ,$$

where we have made use of the well-known identity[5]

$$\tan z = \sum_{n=0}^{\infty} \frac{2z}{[(2n+1)\pi/2]^2 - z^2} .$$

In view of (53′), (52) becomes

(52′)
$$py(p) = \frac{\displaystyle\frac{K_1}{\sqrt{k_1}}\, p^{1/2} \sum_{n=0}^{\infty} \frac{E_n p}{k_1\lambda_n^2 + p} - \int_0^\infty f_2\left(\sqrt{\frac{k_2}{p}}\,\eta\right) e^{-\eta}\, d\eta}{\displaystyle\tanh\sqrt{\frac{a^2 p}{k_1}} + \sqrt{\frac{k_2}{k_1}}\,\frac{K_1}{K_2}} = \frac{Y(p)}{Z(p)} .$$

[5] K. Knopp, *Theorie und Anwendung der unendlichen Reihen*, 1922, Chap. 12.

In the language of the operational calculus, the expression $Y(p)/Z(p)$ is the "operational solution" corresponding to the function $\varphi(t)$ defined by (37). As is well known, the formula for inversion of (37) is

$$\varphi(t) = \frac{Y(0)}{Z(0)} + \sum \frac{Y(p_i)}{p_i Z'(p_i)} e^{p_i t}, \tag{54}$$

where the summation is extended over the roots of the transcendental equation

$$Z(p) = \tanh \sqrt{\frac{a^2 p}{k_1}} + \frac{K_1}{K_2} \sqrt{\frac{k_2}{k_1}} = 0, \tag{55}$$

provided the latter has no zero or repeated roots, conditions which are clearly satisfied.

We evidently have $Y(0) = 0$. Furthermore, $Z(p) > 0$ for $p > 0$, as is at once apparent from the original form of $Z(p)$ (equation 52). Accordingly, it is seen from (55) that $\varphi(t) \rightarrow 0$ as $t \rightarrow \infty$ which is, of course, obvious from physical considerations. Formulae (54), (17) and (32) contain the complete solution of our problem.

Yeshiva College, New York City.

ON PROPERTIES OF REGIONS WHICH PERSIST IN THE SUBREGIONS BOUNDED BY LEVEL CURVES OF THE GREEN'S FUNCTION

By Lester R. Ford

1. Let the unit circle $|z| < 1$, which we shall call Q, be mapped by

$$w = f(z), \qquad f(0) = 0,$$

in a one-to-one and conformal manner on a region S in the w-plane. Let S_r be the map of $|z| < r < 1$, the circle Q_r.

The regions S_r have been extensively cultivated. It is known that if S is a convex region, then S_r is convex also. The simplest proof of this is due to Radó.[1] If S is star-shaped with respect to the origin, the like is true of S_r.[2]

These results raise the question of more general properties of S which hold in the subregions S_r. A generalization which includes the properties just mentioned is given here. The method of proof is suggested by Radó's paper.

2. **The property** T. Let $T(w_1, w_2, \cdots, w_n)$ be analytic in $w_1, w_2, \cdots, w_n$ when these variables range over S, and let $T(0, 0, \cdots, 0) = 0$. We shall say that S has the property T if when $w_1, w_2, \cdots, w_n$ lie in S so also does w_0, where

$$w_0 = T(w_1, w_2, \cdots, w_n).$$

As an example, S is convex if any point w_0 on the line segment joining any two points w_1 and w_2 of S is in S:

$$w_0 = T(w_1, w_2) = tw_1 + (1 - t)w_2, \qquad 0 < t < 1.$$

Again, S is star-shaped from the origin if any point w_0 on the line segment joining the origin to any point w_1 of S is in S:

$$w_0 = T(w_1) = tw_1, \qquad 0 < t < 1.$$

Some of the simplest functions T define properties that have not been studied and lead to interesting regions. Consider $T(w_1) = \frac{1}{2}w_1$. S has the property T if the midpoint of the line segment joining the origin to any point of S lies in S. An instructive region with this property is what remains of the unit circle

Received, February 11, 1935; presented to the American Mathematical Society, July 30, 1934.

[1] T. Radó, *Bemerkung über die konformen Abbildungen konvexer Gebiete*, Math. Ann., vol. 102 (1930), pp. 428–429. The theorem goes back to E. Study, *Konforme Abbildung einfach-zusammenhängender Bereiche*, Leipzig, 1913, p. 110.

[2] W. Seidel, *Über die Ränderzuordnung bei konformen Abbildung*, Math. Ann., vol. 104 (1931), p. 204.

$|w| < 1$ after the removal of all points on lines drawn parallel to the imaginary axis and in the first quadrant from the points $1/2, 1/4, \cdots, 1/2^k$.

3. **Persistence of the property** T. We now establish the principal result of the paper.

THEOREM. *If S has the property T so also has S_r.*

We are to prove that if $w_1, \cdots, w_n$ lie in S_r then w_0, which we know to be in S, is in S_r; or, what amounts to the same thing, if the corresponding points $z_1, \cdots, z_n$ lie in Q_r so also does z_0.

Let z' lie in Q_r and have a greater absolute value than any of the quantities $z_1, \cdots, z_n$; that is,

$$|z_k| < |z'| < r.$$

Let φ be the function inverse to f, so that

$$z = \varphi(w)$$

maps S on Q and S_r on Q_r. We form the function

$$F(z) = \varphi\{T[f(z_1z/z'), \cdots, f(z_nz/z')]\}.$$

We show that for z in Q $F(z)$ is analytic and has a value lying in Q. If z is in Q, so also is z_kz/z', and $f(z_kz/z')$ is in S. Then T is also in S, since S has the property T; and hence $\varphi(T)$ is in Q. We have then

$$|F(z)| < 1.$$

It is clear that $F(z)$ is analytic in Q, since in the preceding sequence we have always analytic functions of analytic functions. Finally, we have

$$F(0) = \varphi\{T[f(0), \cdots, f(0)]\} = \varphi\{T(0, \cdots, 0)\} = \varphi(0) = 0.$$

We now have all the conditions necessary for the application of Schwarz's lemma; whence we have for z in Q

$$|F(z)| \leqslant |z|.$$

In particular

$$|F(z')| \leqslant |z'| < r;$$

that is,

$$\begin{aligned} |\varphi\{T[f(z_1), \cdots, f(z_n)]\}| &= |\varphi\{T(w_1, \cdots, w_n)\}| \\ &= |\varphi\{w_0\}| = |z_0| < r. \end{aligned}$$

Hence z_0 is in Q_r; and the proof is complete.

A large number of applications of this theorem, using various functions T, come readily to mind. The author has not attempted a systematic study of properties of this type.

THE RICE INSTITUTE.

GENERALIZATION OF TWO THEOREMS OF HARDY AND LITTLEWOOD ON POWER SERIES

By Otto Szász

1. Let

$$P(x) = \sum_0^\infty a_\nu x^\nu = a_0 + a_1 x + a_2 x^2 + \cdots$$

be a power series convergent for $0 \leqq x < 1$. It was proved by J. E. Littlewood in 1911 that the existence of

$$\lim_{x \to 1} P(x) = s \tag{1}$$

and the condition

$$na_n = O(1), \qquad n \to \infty \tag{2}$$

imply the convergence of the series $\sum_0^\infty a_\nu$ to s. In 1914 Hardy and Littlewood (1)[1] succeeded in replacing the condition (2) by the "one-sided" condition:

$$na_n \geqq -K, \qquad K \geqq 0, \qquad n = 1, 2, 3, \cdots . \tag{3}$$

Finally R. Schmidt in 1925 replaced this condition by the more general one, viz.,

$$\liminf (a_{n+1} + \cdots + a_m) \geqq 0 \quad \text{as} \quad m/n \to 1 . \tag{4}$$

Another generalization of (2) is:

$$\sum_1^n \nu^p \mid a_\nu \mid^p = O(n) , \qquad n \to \infty , \qquad p > 1 ; \tag{5}$$

it can be derived from (4) as Hardy and Littlewood have observed. A direct proof that (1) and (5) imply convergence was given in my paper (2). A similar generalization of (3) is

$$\sum_1^n \nu^p (\mid a_\nu \mid - a_\nu)^p = O(n) , \qquad n \to \infty , \qquad p > 1 . \tag{6}$$

On writing (3) in the equivalent form

$$n(\mid a_n \mid - a_n) \leqq 2K , \tag{3'}$$

we see that (3) implies (6); hence we get a generalization of Hardy and Littlewood's theorem if we prove

Received December 26, 1934.

[1] See the list of references at the end of this paper.

THEOREM I. *Conditions* (1) *and* (6) *imply the convergence of the series* $\sum_0^\infty a_\nu$ *to* s.

To prove this, it is sufficient to show that (6) implies (4), or, what is the same,

$$\limsup\,(-a_{n+1} - \cdots - a_m) \leqq 0 \quad \text{as} \quad m/n \to 1\,, \tag{4'}$$

and then to apply R. Schmidt's theorem.

Now $-2a_\nu \leqq |a_\nu| - a_\nu$, hence

$$-\sum_{n+1}^m a_\nu \leqq \sum_{n+1}^m (|a_\nu| - a_\nu) = \sum_{n+1}^m \nu(|a_\nu| - a_\nu)\cdot\frac{1}{\nu}\,,$$

and by Cauchy-Hölder's inequality

$$\sum_{n+1}^m (|a_\nu| - a_\nu) \leqq \Big(\sum_{n+1}^m \nu^p(|a_\nu| - a_\nu)^p\Big)^{1/p}\Big(\sum_{n+1}^m \nu^{-p'}\Big)^{1/p'}, \qquad p' = p/(p-1)\,.$$

Thus by (6)

$$\sum_{n+1}^m (|a_\nu| - a_\nu) \leqq \frac{(m-n)^{1/p'}}{n+1}\,O(m^{1/p})\,.$$

On putting here $m \leqq (1+\delta)n$, $0 < \delta < 1$ we get

$$-\sum_{n+1}^m a_\nu \leqq \frac{\delta^{1/p'}\,n^{1/p'}}{n+1}\,O(n^{1/p}) = O(\delta^{1/p'}) \to 0 \quad \text{as} \quad \delta \to 0\,,$$

from which (4′) immediately follows.

In the following lines I give a simple direct proof of Theorem I; it is based on some auxiliary theorems which may be of interest in themselves.

A corresponding theorem for Dirichlet's series can be proved in a similar manner, generalizing a theorem given in my paper (3).

The generality of condition (6) increases as p decreases, since

$$\Big(\frac{1}{n}\sum_1^n \alpha_\nu^p\Big)^q \leqq \Big(\frac{1}{n}\sum_1^n \alpha_\nu^q\Big)^p, \qquad \alpha_\nu > 0\,, \qquad 0 < p < q\,. \tag{7}$$

2. In my paper (2) I proved

THEOREM II. *If* (1) *holds and if*

$$\sum_1^n \nu a_\nu \geqq -Kn\,, \qquad n = 1, 2, 3, \cdots, \tag{8}$$

then $\sum_0^\infty a_\nu$ *is summable* $(C, 1)$ *to the sum* s.

Here is a simpler proof of this theorem. On setting

$$\sum_0^n a_\nu = A_n\,, \qquad \sum_0^n A_\nu = A_n^{(1)}\,, \qquad \sum_1^n \nu a_\nu = (n+1)\,A_n - A_n^{(1)} = v_n\,,$$

we have by (1), as $x \to 1$,

$$P(x)/(1-x) = \sum_0^\infty A_\nu x^\nu \sim s/(1-x)\,,$$

$$P(x)/(1-x)^2 = \sum_0^\infty A_\nu^{(1)} x^\nu \sim s/(1-x)^2\,,$$

$$\text{(9)} \quad \int_0^x P(t)/(1-t)^2\,dt = \sum_0^\infty \frac{1}{\nu+1} A_\nu^{(1)} x^{\nu+1} \sim s \int_0^x dt/(1-t)^2 = s/(1-x)\,.$$

Hence

$$\sum_1^\infty \left(A_\nu - \frac{1}{\nu+1} A_\nu^{(1)}\right) x^\nu = \sum_1^\infty \frac{1}{\nu+1} v_\nu x^\nu = o\,(1/(1-x))\,,$$

or

$$\sum_1^\infty \left(K + \frac{1}{\nu+1} v_\nu\right) x^\nu \sim K/(1-x)\,, \qquad x \to 1\,.$$

Using a well known theorem of Hardy and Littlewood, of which Karamata gave an elementary proof, we derive from this and (8),

$$\sum_1^n \left(K + \frac{1}{\nu+1} v_\nu\right) \sim Kn\,, \qquad n \to \infty\,,$$

or

$$\text{(10)} \qquad \sum_1^\infty \frac{1}{\nu+1} v_\nu = o(n)\,.$$

By (9)

$$\frac{1-x}{x} \int_0^x P(t)/(1-t)^2\,dt = a_0 + \sum_1^\infty \left(\frac{1}{\nu+1} A_\nu^{(1)} - \frac{1}{\nu} A_{\nu-1}^{(1)}\right) x^\nu \to s\,.$$

On setting here

$$\frac{1}{\nu+1} A_\nu^{(1)} - \frac{1}{\nu} A_{\nu-1}^{(1)} = \frac{1}{\nu}\,(A_\nu^{(1)} - A_{\nu-1}^{(1)}) - \frac{1}{\nu(\nu+1)} A_\nu^{(1)} = \frac{v_\nu}{\nu(\nu+1)} = u_\nu\,,$$

we have by (10),

$$\sum_1^n \nu u_\nu = o(n)\,,$$

whence, by Tauber's original theorem,

$$\frac{1}{n+1} A_n^{(1)} \to s, \qquad n \to \infty .$$

I am going to prove now

THEOREM III. *If* $\sum_0^\infty a_\nu$ *is summable* $(C, 1)$ *to* s *and if* (6) *holds, then* $\sum_0^\infty a_\nu$ *converges to* s.

We use the identity

$$A_n - \frac{A_{n+\nu}^{(1)}}{n+\nu+1} = \frac{n}{\nu+1}\left(\frac{1}{n+\nu+1} A_{n+\nu}^{(1)} - \frac{1}{n} A_{n-1}^{(1)}\right) - \frac{1}{\nu+1} \sum_{k=1}^{\nu} (\nu - k + 1)\, a_{n+k}. \tag{11}$$

Now

$$-\sum_{k=1}^{\nu}(\nu - k + 1)\, a_{n+k} \leqq \sum_{k=1}^{\nu}(\nu - k + 1)\,(\,|\,a_{n+k}\,| - a_{n+k}) \leqq \nu \sum_{k=1}^{\nu}(\,|\,a_{n+k}\,| - a_{n+k}), \tag{12}$$

and using (7) and (6),

$$\sum_{n+1}^{n+\nu}(\,|\,a_k\,| - a_k) \leqq \nu^{1-1/p}\left[\sum_{n+1}^{n+\nu}(\,|\,a_k\,| - a_k)^p\right]^{\frac{1}{p}} \leqq \frac{1}{n+1}\,\nu^{1-1/p}\left[\sum_{n+1}^{n+\nu} k^p\,(\,|\,a_k\,| - a_k)^p\right]^{\frac{1}{p}} = O(\nu^{1-1/p}(n+\nu)^{1/p}(n+1)^{-1}) . \tag{13}$$

Being given an arbitrary $\epsilon > 0$, by assumption, we have

$$\left|\frac{1}{n+\nu+1} A_{n+\nu}^{(1)} - \frac{1}{n} A_{n-1}^{(1)}\right| < \epsilon^2 \text{ for } n > N(\epsilon), \qquad \nu = 1, 2, 3, \cdots ; \tag{14}$$

choose $\nu = [n\epsilon]$ and $n > \epsilon^{-1}$, then by (11)–(13)

$$\limsup_{n\to\infty} A_n - s \leqq \epsilon^{-1}\epsilon^2 + O(\epsilon^{1-1/p}) ,$$

hence

$$\limsup A_n \leqq s.$$

Similarly from

$$A_n - \frac{1}{n-\nu} A^{(1)}_{n-\nu-1} = \frac{n+1}{\nu+1}\left(\frac{A^{(1)}_n}{n+1} - \frac{A^{(1)}_{n-\nu-1}}{n-\nu}\right) + \frac{1}{\nu+1}\sum_{k=1}^{\nu} (\nu - k + 1) a_{n-k+1} \tag{15}$$

it follows

$$\liminf_{n\to\infty} A_n \geqq s,$$

and this proves Theorem III.

Since, by (7) and (6),

$$-\sum_1^n \nu a_\nu \leqq \sum_1^n \nu(|a_\nu| - a_\nu) \leqq n^{1-1/p}\left(\sum_1^n \nu^p(|a_\nu| - a_\nu)^p\right)^{1/p} = O(n),$$

Theorem II asserts that $\sum_0^\infty a_\nu$ is summable $(C, 1)$ to s, and by Theorem III, $\sum_0^\infty a_\nu$ converges to s; this proves Theorem I.

3. It may be observed that (6) is equivalent to the following condition: there exists a $g > 1$ such that

$$\sum_n^{[gn]} (|a_\nu| - a_\nu)^p = O(n^{1-p}), \qquad n \to \infty. \tag{6'}$$

For from (6) it follows

$$\sum_n^{2n} (|a_\nu| - a_\nu)^p \leqq n^{-p}\sum_n^{2n} \nu^p(|a_\nu| - a_\nu)^p = O(n^{1-p}).$$

Conversely on putting $n_\nu = [ng^{-\nu}]$, $\nu = 0, 1, 2, \cdots$,

$$\sum_1^n \nu^p(|a_\nu| - a_\nu)^p = \sum_\nu \sum_{1+n_\nu}^{n_{\nu-1}} k^p(|a_k| - a_k)^p \leqq \sum_\nu n^p_{\nu-1} \sum_{1+n_\nu}^{n_{\nu-1}} (|a_k| - a_k)^p,$$

and using (6′),

$$\sum_1^n \nu^p(|a_\nu| - a_\nu)^p = O\left(\sum_\nu n_\nu\right) = O\left(n\sum_0^\infty g^{-\nu}\right) = O(n).$$

A generalization of (6) and (6′) is

$$\sum_{n+1}^{[n(1+\delta)]} (|a_\nu| - a_\nu) = O(\omega(\delta)), \qquad \omega(\delta) \to 0 \text{ as } \delta \to 0. \tag{16}$$

For from (6) it follows

$$\sum_{n+1}^{n(1+\delta)]} (|a_\nu| - a_\nu) \leqq \frac{1}{n+1} \sum \nu(|a_\nu| - a_\nu)$$

$$\leqq \frac{1}{n+1} \left(\sum \nu^p(|a_\nu| - a_\nu)^p\right)^{1/p} (n\delta)^{1-1/p} \leqq (1+\delta)^{1/p}\delta^{1-1/p} O(1) \to 0 \text{ as } \delta \to 0.$$

The same argument as in §2 leads to the more general

THEOREM IV. *Conditions* (1) *and* (16) *imply the convergence of* $\sum_0^\infty a_\nu$ *to* s.

In the first place, on setting $n_\nu = [n(1+\delta)^{-\nu}]$ we have

$$-\sum_1^n \nu a_\nu \leqq \sum_1^n \nu(|a_\nu| - a_\nu) = \sum_\nu \sum_{1+n_\nu}^{n_{\nu-1}} k(|a_k| - a_k) \leqq \sum_\nu n_{\nu-1} O(\omega(\delta)) = O(n);$$

hence by Theorem II, $\sum_0^\infty a_\nu$ is summable $(C, 1)$ to s. Now

$$-\sum_{k=1}^{\nu} (\nu - k + 1) a_{n+k} \leqq \sum_1^\nu (\nu - k + 1)(|a_{n+k}| - a_{n+k}) \leqq \nu \sum_{n+1}^{n+\nu} (|a_k| - a_k)$$

$$= \nu O(\omega(\epsilon)), \qquad \nu = [n\epsilon], \quad n > \epsilon^{-1};$$

hence by (11), (12) and (14)

$$\limsup_{n\to\infty} A_n \leqq s.$$

Similarly from (15) it follows

$$\liminf_{n\to\infty} A_n \geqq s;$$

this proves the theorem.

4. Hardy and Littlewood proved that (1) and the convergence of the series $\sum \nu^\rho |a_\nu|^{\rho+1}$, $\rho > 0$, imply the convergence of $\sum_0^\infty a_\nu$; the case $\rho = 1$ was treated by Fejér. I prove the following generalization:

THEOREM V. *The conditions* (1) *and*

$$\sum_1^\infty \nu^\rho(|a_\nu| - a_\nu)^{\rho+1} < \infty, \qquad \rho > 0.$$

imply the convergence of $\sum_0^\infty a_\nu$ *to* s.

This follows immediately from Theorem I; for

$$\sum_n^{2n} (|a_\nu| - a_\nu)^{\rho+1} \leqq n^{-\rho} \sum_n^{2n} \nu^\rho(|a_\nu| - a_\nu)^{\rho+1} = o(n^{-\rho}),$$

hence (6′) is satisfied with $p = \rho + 1 > 1$. At the same time it is clear that Theorem V is of the "o-type."

REFERENCES

1. G. H. Hardy and J. E. Littlewood, *Tauberian theorems concerning power series and Dirichlet series whose coefficients are positive*, Proc. London Math. Soc., (2), vol. 13 (1914), pp. 174–191.
2. O. Szász, *Verallgemeinerung eines Littlewoodschen Satzes über Potenzreihen*, Journ. London Math. Soc., vol. 3 (1928), pp. 254–262.
3. O. Szász, *Über Dirichletsche Reihen an der Konvergenzgrenze*, Atti Congresso Internaz. Bologna, vol. III, pp. 269–276.

Massachusetts Institute of Technology and Brown University.

LINEAR ALGEBRAS WITH ASSOCIATIVITY NOT ASSUMED

By L. E. Dickson

1. The complete structure of linear associative algebras was known to depend upon the division algebras. When the reference field F is an algebraic field, H. Hasse has recently proved that every normal[1] division algebra is cyclic. This perfection of the theory of associative algebras justifies attention to non-associative algebras.

Known examples of non-associative division algebras are Cayley's algebra of order 8, and the writer's[2] commutative algebras of orders 3 and $2n$ (§15). Many new division algebras of order 4 are given here by Theorems 2 and 3.

In §§7–11 we determine all types of algebras of order 3 having a principal unit (or modulus) denoted by 1. Except for special values of the parameters, these algebras are simple. It is known that every associative simple algebra of order 3 is a division algebra.

Thus the structure theorems for associative algebras fail in general for non-associative algebras. Similarly for other properties. Consider the algebra A of order 4 with $e_1^2 = e_2$, $e_2e_1 = e_3$, and all further $e_ie_j = 0$. For $X = x_0 + \Sigma x_ie_i$, evidently $(X - x_0)(X - x_0)^2 = 0$, so that A has the left rank 3. But its right rank is 4 since $1, e_1, e_1^2 = e_2, e_1^2\, e_1 = e_3$ are linearly independent. Algebra (37) with 13 parameters also has left and right ranks 3 and 4.

Part I. Rank 2

2. In case the field F has a modulus p, assume that $p \neq 2$.

Lemma 1. *If* $1, u, v$ *are linearly independent with respect to* F, *and* $u^2 = a + cu$, $v^2 = b + dv$, *then*

$$uv + vu = du + cv + f, \qquad f \text{ in } F.$$

Write t for $a + b + cu + dv$. Then

$$(u + v)^2 = t + uv + vu = r + s(u + v),$$

$$(u - v)^2 = t - uv - vu = R + S(u - v).$$

Addition yields $s + S = 2c$, $s - S = 2d$. Subtraction gives Lemma 1.

Received March 25, 1935.

[1] If only the numbers of F are commutative with every element.

[2] *Linear Algebras*, Cambridge Tract No. 16, p. 14, p. 69 (p. 17 for the characteristic equations); *On triple algebras and ternary cubic forms*, Bull. Amer. Math. Soc., vol. 14 (1907–8), pp. 160–169, p. 169; *Linear algebras in which division is always uniquely possible*, Trans. Amer. Math. Soc., vol. 7 (1906), pp. 370–390; *On commutative linear algebras in which division is always uniquely possible*, ibid., pp. 514–522.

3. After adding scalars (numbers of F) to the initial units, we obtain a basis $1, e_1, \cdots, e_n$, where $e_i^2 = a_i$. By Lemma 1,

$$e_ie_j + e_je_i = b_{ij} \qquad (i \neq j).$$

We get e_je_i from the latter and

$$e_ie_j = c_{ij} + \sum_{k=1}^{n} c_{ijk}e_k \qquad (i < j).$$

Using summations for $i = 1, \cdots, n$, let

$$X = \xi + \Sigma x_ie_i, \qquad Y = \eta + \Sigma y_ie_i, \qquad XY = P_0 + \Sigma P_ie_i.$$

Let Σ' denote summation for $i, j = 1, \cdots, n$ with $i < j$. Then

$$P_0 = \xi\eta + \Sigma x_iy_ia_i + \Sigma' b_{ij}x_jy_i + \Sigma'(x_iy_j - x_jy_i)c_{ij},$$

$$P_k = \xi y_k + \eta x_k + \Sigma'(x_iy_j - x_jy_i)c_{ijk} \qquad (k = 1, \cdots, n).$$

The right determinant $\Delta(X)$ is defined to be the determinant of the coefficients of $\eta, y_1, \cdots, y_n$ in $P_0, P_1, \cdots, P_n$. Multiply the first column by $-\xi$ and the $(j + 1)$-st column by x_j (for $j = 1, \cdots, n$) and add. The sum is the value of P_k for $\eta = -\xi, y_j = x_j$. For $k \geqq 1$, P_k evidently vanishes. Since Y becomes $-X'$, where X' is the conjugate of X, P_0 becomes the negative of the norm of X.

THEOREM 1. *The right determinant $\Delta(X)$ and left determinant $\Delta'(X)$ of any algebra of rank 2 are both divisible by the norm $N(X)$.*

The quotient is obtained at once from the cofactor of any element in the first row (cf. §5).

4. **Division algebras of rank 2.** We assume that if any product of two factors is zero, then one factor is zero.

LEMMA 2. *If 1, u, v are linearly independent in F, then 1, u, v, uv are linearly independent in F.*

We can choose m and n in F such that

$$U = u - m, \qquad V = v - m, \qquad U^2 = a, \qquad V^2 = k.$$

By Lemma 1, $UV + VU = 2c$. Take $e_1 = U$, $e_2 = V - ca^{-1}U$. Then

$$e_1^2 = a, \qquad e_2^2 = b, \qquad e_1e_2 + e_2e_1 = 0.$$

If $a = k^2$, k in F, then $(e_1 - k)(e_1 + k) = 0$, $e_1 = \pm k$, whereas 1 and e_1 are linearly independent in F. Hence

$$(1) \qquad a \neq \text{square in } F, \qquad b \neq \text{square in } F.$$

If 1, u, v, uv are linearly dependent in F, the same is true of 1, U, V, UV and of 1, e_1, e_2, e_1e_2. Hence $e_1e_2 = r + se_1 + te_2$, where r, s, t are in F. Then

$$(e_1 - t)\left(e_2 + \frac{r + st}{t^2 - a}e_1 + \frac{rt + as}{t^2 - a}\right) = 0\,,$$

whereas neither factor is zero. This proves Lemma 2.

Since we may take e_1e_2 as the fourth basal element e_3,

$$(2) \qquad e_1^2 = a, \qquad e_2^2 = b, \qquad e_1e_2 = e_3, \qquad e_2e_1 = -e_3\,.$$

Write $e_i^2 = c_i + 2m_ie_i$ for $i \geqq 3$. Apply Lemma 1 with $(u, v) = (e_1, e_i)$, (e_2, e_i), (e_i, e_j) in turn. We get

$$(3) \qquad \begin{aligned} &e_1e_i + e_ie_1 = g_i + 2m_ie_1, \qquad e_2e_i + e_ie_2 = h_i + 2m_ie_2 \qquad (i \geqq 3)\,, \\ &e_ie_j + e_je_i = t_{ij} + 2m_ie_j + 2m_je_i \qquad (i, j \geqq 3, i \neq j)\,. \end{aligned}$$

Write $x = \xi + \Sigma x_ie_i$. Squaring $x - \xi$, we get

$$x^2 = 2x\left(\xi + \sum_{j=3}^{n} x_jm_j\right) - N\,,$$

$$N = \xi^2 + 2\xi\Sigma x_jm_j - ax_1^2 - bx_2^2 - \Sigma c_ix_i^2 - \Sigma g_jx_1x_j - \Sigma h_jx_2x_j - \Sigma' t_{ij}x_ix_j,$$

where the summations are for $i \geqq 3, j \geqq 3$, while $i < j$ in Σ'. Thus (2) and (3) imply that every element x of the algebra satisfies a quadratic equation. Evidently N is the norm xx' of x. Also,

$$(4) \qquad x^2 \text{ is scalar if and only if } \xi = -\sum_{j=3}^{n} x_jm_j\,.$$

We may choose a basis satisfying (2) and e_3^2 = scalar. Proof is needed only when $m_3 \neq 0$. Use the basis

$$1, \qquad e_1, \qquad E_2 = x = \xi + \Sigma x_ie_i, \qquad E_3 = e_1E_2, \qquad \cdots,$$

where (4) holds. Then

$$e_1E_2 + e_2E_1 = 2ax_1 + \sum_{i=3}^{n} g_ix_i$$

becomes zero by choice of x_1. Let

$$e_1e_i = k_i + \sum_{j=1}^{n} c_{ij}e_j \qquad (i \geqq 3)\,.$$

Then

$$E_3 = s + \xi e_1 + x_2e_3 + \sum_{j=1}^{n}\left(\sum_{i=3}^{n} x_ic_{ij}\right)e_j\,, \quad s = ax_1 + \sum_{i=3}^{n} x_ik_i.$$

By (4), E_3^2 is scalar if

$$s + m_3(x_2 + \sum_{i=3}^{n} x_ic_{i3}) + \sum_{j=4}^{n} m_j\left(\sum_{i=3}^{n} x_ic_{ij}\right) = 0\,.$$

This and the former condition hold if, for example,

$$x_i = 0\ (i \geqq 4)\,, \qquad x_3 = 2, \qquad x_1 = -g_3/a, \qquad m_3x_2 = g_3 - 2k_3 - 2\sum_{j=3}^{n} m_jc_{3j}\,.$$

LEMMA 3. *There exists a basis satisfying* (2) *and* (3) *with* $m_3 = 0$.

5. **Case** $n = 3$. Employ (2) and

$$(5)\quad \begin{array}{lll} e_1e_3 = c_0 + c_1e_1 + c_2e_2 + c_3e_3, & & e_2e_3 = d_0 + d_1e_2 + d_2e_2 + d_3e_3, \\ e_3^2 = f, & e_3e_1 = g - e_1e_3, & e_3e_2 = h - e_2e_3, \end{array}$$

which follow from (3). As in Section 3, $\Delta(x)$ is

$$\begin{vmatrix} \xi & ax_1 + (g - c_0)x_3 & bx_2 + (h - d_0)x_3 & c_0x_1 + d_0x_2 + fx_3 \\ x_1 & \xi - c_1x_3 & -d_1x_3 & c_1x_1 + d_1x_2 \\ x_2 & -c_2x_3 & \xi - d_2x_3 & c_2x_1 + d_2x_2 \\ x_3 & -x_2 - c_3x_3 & x_1 - d_3x_3 & c_3x_1 + d_3x_2 + \xi \end{vmatrix},$$

which is the product of

$$(6)\quad -N(x) = -\xi^2 + ax_1^2 + bx_2^2 + fx_3^2 + gx_1x_3 + hx_2x_3$$

and the quotient[3] by $-x_3$ of the minor obtained by deleting the first row and last column. This quotient is

$$(7)\quad \begin{array}{l} -\xi^2 - c_3\xi x_1 - d_3\xi x_2 + (c_1 + d_2)\xi x_3 + c_2x_1^2 + (d_2 - c_1)x_1x_2 \\ \qquad -d_1x_2^2 + (c_3d_2 - c_2d_3)x_1x_3 + (c_1d_3 - c_3d_1)x_2x_3 + (c_2d_1 - d_1d_2)x_3^2. \end{array}$$

Similarly, the left determinant $\Delta'(x)$ is the product of (6) by the function derived from (7) by changing the sign of ξ.

THEOREM 2. *The algebra R over F defined by* (2) *and* (5) *is a division algebra if and only if neither* (6) *nor* (7) *is zero for values of* ξ, x_1, x_2, x_3 *in the field not all zero. The forms* (6) *and* (7) *are identical if and only if R is the algebra S defined by* (2) *and*

$$(8)\quad \begin{array}{lll} e_3^2 = -ab, & e_1e_3 = c_0 + ae_2, & e_3e_1 = -e_1e_3, \\ & e_2e_3 = d_0 - be_1, & e_3e_2 = -e_2e_3. \end{array}$$

In R the product of the norms of any two elements is the norm of their product if and only if R is the algebra Q defined by (2) *and* (8) *with* $c_0 = d_0 = 0$, *whence Q is the generalized quaternion algebra and obeys the associative law.*

In the products of (6) and (7) by -1 and -4, respectively, we complete the square on ξ, then on x_1, and finally on x_2. We obtain forms having only square terms with the coefficients

$$(6')\quad 1,\ -a,\ -b,\ -f + g^2/(4a) + h^2/(4b)\ ;$$

$$(7')\quad 1,\ -A,\ -B,\ D + C^2/B,$$

[3] Or the quotient by x_2 of the minor obtained by deleting the first row and third column, etc.

where

$$A = c_3^2 + 4c_2, \qquad B = d_3^2 - 4d_1 - U^2/A, \qquad U = 2(c_1 - d_2) - c_3d_3,$$

$$W = c_1c_3 + 2c_2d_3 - c_3d_2, \qquad C = 2c_3d_1 - c_1d_3 + d_2d_3 + UW/A,$$

$$D = 4(c_1d_2 - c_2d_1) - (c_1 + d_2)^2 + W^2/A.$$

When F is the field of all real numbers, the condition for a division algebra is that the numbers (6′) and (7′) be all positive.

When F is any field, $x^2 - ay^2 - bz^2 + mw^2$ is not zero for values of $x, \cdots, w$ in F not all zero if and only if a is not a square in F, b is not represented by $x^2 - ay^2$, nor $-m$ by $x^2 - ay^2 - bz^2$, for values in F of the variables. The final condition is redundant[4] if $m = ab$.

The case $c_3 = d_3 = c_1 = d_2 = g = h = 0, f = -ab$, yields the

COROLLARY. *The algebra defined by* (2) *and*

$$e_1e_3 = c_0 + ce_2, \quad e_2e_3 = d_0 - de_1, \quad e_3^2 = -ab, \quad e_3e_1 = -e_1e_3, \quad e_3e_2 = -e_2e_3$$

is a division algebra if and only if a and c are not squares in F, b is not represented by $x^2 - ay^2$, and d is not represented by $x^2 - cy^2$. When $c = a$, $d = b$, this algebra is S of Theorem 2.

6. **Normalization.** Employ a new basis of R:

$$(9) \qquad 1, \quad E_1, \quad E_2, \quad E_3 = E_1E_2, \quad E_1^2 \text{ and } E_2^2 \text{ scalar.}$$

By (4)

$$(10) \qquad E_1 = \Sigma x_ie_i, \qquad E_2 = \Sigma y_ie_i \qquad (i = 1, 2, 3).$$

By Lemma 2, (9) is a basis if and only if the x_i are not proportional to the y_i. The condition for $E_1E_2 + E_2E_1 = 0$ is $C(x_i, y_i) = 0$, viz.,

$$(11) \quad 2t + g(x_1y_3 + x_3y_1) + h(x_2y_3 + x_3y_2) = 0, \quad t = ax_1y_1 + bx_2y_2 + fx_3y_3.$$

We get $E_3 = f_0 + \Sigma f_ie_i$, where

$$f_0 = t + c_0x_1y_3 + (g - c_0)x_3y_1 + d_0x_2y_3 + (h - d_0)x_3y_2,$$

$$(12) \qquad f_i = c_iL + d_iM \quad (i = 1, 2), \qquad f_3 = c_3L + d_3M + N,$$

$$L = x_1y_3 - x_3y_1, \qquad M = x_2y_3 - x_3y_2, \qquad N = x_1y_2 - x_2y_1.$$

We require that E_3^2 be scalar, viz., $f_0 = 0$. Next $C(x_i, f_i) = 0$ is the condition for $E_1E_3 + E_3E_1 = 0$. We now have three linear homogeneous equations in y_1, y_2, y_3. The determinant D of their coefficients is a quartic form in the x_i. Assume that $D = 0$ has solutions not all zero in F. Then our three equations hold if the y_i are the cofactors of the elements of the third row of D. For these values (quadratic functions of the x_i) of the y_i, the condition $C(y_i, f_i) = 0$

[4] L. E. Dickson, *Algebren und ihre Zahlentheorie*, Zürich, 1927, p. 47.

for $E_2E_3 + E_3E_2 = 0$ becomes an equation in the x_i of degree 5. This has solutions in common with $D = 0$, but not necessarily in F. Hence in general we may normalize algebra R so that $g = h = 0$.

Consider the normalization of algebra S defined by (2) and (8). Here (11) and $f_0 = 0$ become

$$(13) \qquad ax_1y_1 + bx_2y_2 - abx_3y_3 = 0, \qquad c_0L + d_0M = 0.$$

When (13) hold, the new basis (9) satisfies equations (2) and (8) written in capital letters if and only if

$$(14) \qquad A = ax_1^2 + bx_2^2 - abx_3^2, \qquad B = ay_1^2 + by_2^2 - aby_3^2,$$

$$(15) \qquad AB + ab(aL^2 + bM^2 - N^2) = 0,$$

$$(16) \quad C_0 = ab(x_2L - x_1M - x_3N) + c_0(x_1N + bx_3M) + d_0(x_2N - ax_3L),$$

while D_0 is derived from (16) by replacing x_i by y_i. Using the definitions (12) and (14), we find that the left member of (15) reduces to the square of the first sum in (13). Hence we may discard (15).

To make $C_0 = 0$ for solutions of (13), we have three linear homogeneous equations in the y_i; the determinant of their coefficients is a quartic form in the x_i, which must vanish.

For the special case $c_0 = d_0 = 0$, C_0 and D_0 are identically zero by (12). Hence *every normalization of a generalized quaternion algebra is made by the new units* (10) *and* $E_3 = E_1E_2$, *where the* x_i *are not proportional to the* y_i *and satisfy the single equation* (13_1); *the new parameters are* (14).

7. **Triple algebras of rank 2.** By the first two sentences of §3,

$$e_1^2 = a, \qquad e_2^2 = b, \qquad e_1e_2 = c + de_1 + fe_2, \qquad e_2e_1 = g - de_1 - fe_2.$$

If $d = f = 0$, we have (17) with $k = 0$. If d and f are not both zero, we may interchange the e's, if necessary, and take $d \neq 0$. Write

$$E_1 = de_1 + fe_2, \qquad E_2 = d^{-1}e_2,$$

$$E_1E_2 = c + bf/d + E_1, \qquad E_2E_1 = g + bf/d - E_1.$$

Hence every triple algebra of rank 2 is equivalent to[5]

$$(17) \quad e_1^2 = a, \quad e_2^2 = b, \quad e_1e_2 = c + ke_1, \quad e_2e_1 = g - ke_1 \qquad (k = 0 \text{ or } 1).$$

Every transformation of units for which the E_i^2 are scalar is

$$(18) \quad E_1 = re_1 + se_2, \qquad E_2 = te_1 + we_2, \qquad D = rw - st \neq 0.$$

Then $E_1E_2 = kDe_1 +$ scalar. Let (17) hold in capital letters. Then $kDe_1 = KE_1$, whence $kD = Kr$, $Ks = 0$. Hence an algebra (17) with $K = 0$ is not equivalent to one with $k \neq 0$.

[5] If $k = 0$, we may take $c = 0$ unless $a = b = c + g = 0$.

I. $k = 1$. Then $K = 1$, $s = 0$, $w = 1$, $r \neq 0$. Then $E_2E_1 = G - E_1$, $G = r(g + ta)$, $E_1E_2 = C + E_1$, $C = r(c + ta)$.

I_1. $a \neq 0$. We make $G = 0$ by choice of t. Hence we may take $g = 0$. Then $G = 0$ requires $t = 0$, $E_1 = re_1$, $E_2 = e_2$. Then

$$E_1^2 = r^2a, \qquad E_2^2 = b, \qquad E_1E_2 = rc + E_1, \qquad E_2E_1 = -E_1.$$

If $c \neq 0$, take $rc = 1$. Thus

$$e_1^2 = a, \qquad e_2^2 = b, \qquad e_1e_2 = c + e_1, \qquad e_2e_1 = -e_1 \qquad (c = 0 \text{ or } 1), \tag{19}$$

with no further normalization if $c = 1$. But if $c = 0$, we may remove any square factor from a.

I_2. $a = 0$. Then $G = rg$, $C = rc$, $A = r^2a$, $B = b + t(c + g)$. If $c + g \neq 0$, we may take $b = 0$ and a non-vanishing one of c and g equal to 1. If $c + g = 0$, $c \neq 0$, we may take $c = 1$. If $c = g = 0$, we can remove a square factor from a. No further normalization is possible.

For any e in algebra (19), $(1, e)$ is a sub-algebra which is not invariant. Every sub-algebra of order 2 lacking 1 has a basis in which the pairs of coefficients of e_1 and e_2 are not proportional and hence a basis $r + e_1$, $s + e_2$, with

$$c = 2r, \qquad r(s - 1) = 0, \qquad r^2 = a, \qquad s^2 = b.$$

Either $c = a = r = 0$, $b = s^2$, or $c = b = s = 1$, $r = 1/2$, $a = 1/4$. It is invariant by

LEMMA 4. *If S is a sub-algebra of order n of an algebra A of order $n + 1$ and if A has a modulus 1, but 1 is not in S, then S is invariant in A.*

For, $A = (1, S)$, $S^2 \leqq S$. Hence $AS \leqq S$, $SA \leqq S$.

If $c = a = 0$, (19) has the single invariant sub-algebra $[e_1]$ of order 1, otherwise none.

A simple algebra is one having no proper invariant sub-algebra. *Hence* (19) *is simple unless $c = a = 0$ or $c = b = 1$, $a = 1/4$.* Algebra (19) is associative if and only if $a = c = 0$, $b = 1$.

Part II. Triple algebras of rank 3

8. Employ the new basis

$$1, \quad E_1 = a + be_1 + ce_2, \quad E_2 = e + fe_1 + ge_2, \quad D = bg - cf \neq 0. \tag{20}$$

The commutative case shows the existence of a relation

$$E_2E_1 - E_1E_2 = D(e_2e_1 - e_1e_2). \tag{21}$$

We may choose e_1 so that $1, e_1, e_1^2$ are linearly independent in F. Hence we may take

$$\begin{aligned} e_1^2 &= e_2, \qquad e_1e_2 = h + je_1 + ke_2, \\ e_2^2 &= m + ne_1 + pe_2, \qquad e_2e_1 = r + se_1 + te_2. \end{aligned} \tag{22}$$

For any new basis (20) we will have $E_1^2 = E_2$ if

$$(23)\qquad \begin{aligned} e &= a^2 + bc(h + r) + c^2m, \\ f &= 2ab + bc(j + s) + c^2n, \\ g &= 2ac + b^2 + bc(k + t) + c^2p. \end{aligned}$$

Let (22′) denote the multiplication table for E_1, E_2; it is of the form (22) with each letter replaced by the corresponding capital letter. Using also (21), we get

$$(24)\qquad R - H + (S - J)E_1 + (T - K)E_2 = D[r - h + (s - j)e_1 + (t - k)e_2].$$

Solving (20), we get

$$(25)\qquad De_1 = ce - ag + gE_1 - cE_2, \qquad De_2 = af - be - fE_1 + bE_2.$$

Elimination of e_1 and e_2 between (24) and (25) gives

$$(26)\qquad S - J = g(s - j) - f(t - k), \qquad T - K = -c(s - j) + b(t - k).$$

We can always[6] make $S = J$. Proof is needed only when $s \neq j$. By (26), $S = J$ if

$$g = fu, \qquad u = (t - k)/(s - j).$$

By (23), this holds if $2a(c - bu)$ equals a known quadratic form in b, c. This condition is satisfied by choice of a in terms of b and c if $c - bu \neq 0$. Then $D = -f(c - bu) \neq 0$ if $f \neq 0$; we must therefore avoid the values for which f and g in (23) are both zero.

Hence we may take $s = j$. This is not altered if we replace e_1 by $e_1 - k/3$, which yields $K = 0$. Hence we may take

$$(27)\qquad s = j, \qquad k = 0.$$

Apply transformation (20) subject to (23). By (24), $S = J$ if $ft = 0$. There are two cases.

9. Let $t \neq 0$. Then $f = 0$, $D = bg \neq 0$. By (23_2),

$$(28)\qquad -2a = 2cj + c^2n/b.$$

Then

$$S = J = e + gj + gcn/b, \qquad K = a + c(p - j) - c^2n/b.$$

Eliminating a by (28), we see that $K = 0$ only when

$$(29)\qquad c = 0 \qquad \text{or} \qquad 2p - 4j - 3cn/b = 0.$$

[6] But we can make $T = K$ only when a certain invariant is not zero, C. C. MacDuffee, *Invariantive characterizations of linear algebras with the associative law not assumed*, Trans. Amer. Math. Soc., vol. 23 (1922), pp. 135–150, p. 147. In his equation below (27), the final product should be $(b_2 - c_1 - d_1)(c_1 - d_1)(c_2 - d_2)^2 - b_1(c_2 - d_2)^3$, while α_1^3 should be α_2^3.

Now

$$bT = ab + bc(p - j) + b^2t - c^2n$$

reduces by (28) and (29) to

$$T = bt. \tag{30}$$

By choice of b, we may take $T = 1$. To preserve this normalization $t = 1$, we must take $b = 1$. Since

$$N = g^2n/b, \qquad P - 2J = -g(p - 2j), \tag{31}$$

no algebra of one of the following four types is equivalent to one of a different type.

I or II. $n \neq 0$, $p = 2j$, or $n = 0$, $p \neq 2j$. Then $c = 0$ by (29), whence $a = 0$ by (28) and our transformation (20) is the identity.

III. $n \neq 0$, $p \neq 2j$. Then (29) and (28) uniquely determine $c \neq 0$ and a, whence our transformation yields a unique new algebra.

IV. $n = 0$, $p = 2j$. Our transformation involves a single parameter c which is available for further normalization.

10. Let $t = 0$. Then $S = J$, $T = K$ by (26) and (27). After divisions by D, we find that

$$K = 3a + (p + j)c, \qquad J = -3a^2 - 2ac(p + j) + b^2j + bc(n + h + r) + c^2(m - pj),$$

$$P = 6a^2 + 4ac(p + j) + b^2p + bc(2h + 2r - n) + c^2(p^2 + 2pj + 2m).$$

To preserve the normalization $k = 0$, take $K = 0$ and eliminate a. Then J and P become quadratic forms in b, c, while

$$27N = 27nb^3 + 18w^2b^2c - 27wnbc^2 + (2w^3 + 27n^2)c^3,$$

where $w = p - 2j$. The expressions for H, M, R are more complicated, although $R - H = D(r - h)$ by (24). If F is the field of all real numbers, we may take $N = 0$. For a different F, normalization depends on representation by binary forms.

11. Let A be the algebra defined by (22) and (27). Any sub-algebra of order 2 without the modulus 1 is $(u + e_1, v + e_2)$, where

$$v = -u^2, \qquad h = u(v + j), \qquad r - h = tv, \qquad m = nu + v^2 + pv.$$

It is invariant in A by Lemma 4. Elimination of u and v yields two conditions on the constants of A.

Let $W = x + ye_1 + ze_2$ generate an invariant sub-algebra of order 1 of A. We may take $y = 1$ or $y = 0$ and $z = 1$. In the latter case, elimination of $x = -j$ shows that the conditions on A are

$$h = n = 0, \qquad r = -jt, \qquad m = j(j - p).$$

If $y = 1$, the conditions on A are $r = h$, $t = 0$ and the results of eliminating x and z between

$$x = z^2h, \quad z(x + zj) = 1, \quad x(j + zn) = h + zm, \quad z(j + zn) = x + zp.$$

If none of these special sets of conditions holds, A is simple.

Part III. Algebras of order 4 and left rank 3

12. Let xx^2 be a linear combination of 1, x, x^2 for every element x. Let 1, e, e^2 be linearly independent. By choice of r in the field F, $e_1 = e + r$ satisfies the first two equations in (33). Let $e_1e_3 = a_0 + \Sigma a_ie_i$. For

$$E = -a_1 - a_2a_3 - a_2e_1 + e_3, \tag{32}$$

e_1E is linear in E. Take E as a new e_3. Hence

$$\begin{aligned} &e_1^2 = e_2, && e_1e_2 = a + be_1, && e_1e_3 = a_0 + a_3e_3, && e_2e_1 = \Sigma b_ie_i, \\ (33)\quad &e_2^2 = \Sigma c_ie_i, && e_2e_3 = \Sigma d_ie_i, && e_3e_1 = \Sigma f_ie_i, && \\ & && e_3e_2 = \Sigma g_ie_i, && e_3^2 = \Sigma h_ie_i, && \end{aligned}$$

where $e_0 = 1$. Write

$$\begin{aligned} &X = se_1 + te_2 + we_3, \qquad X^2 = K + Le_1 + Me_2 + Pe_3, \\ (34)\quad &XX^2 = \text{scalar} + Re_1 + Te_2 + Ve_3. \end{aligned}$$

The algebra will have left rank 3 if 1 and these three elements are linearly dependent. This is true if and only if

$$\begin{vmatrix} s & L & R \\ t & M & T \\ w & P & V \end{vmatrix} = 0. \tag{35}$$

Since M has the term s^2, while L and P are linear in s, the minors of V and R are not zero identically. We here assume that the minor of T is zero, whence $L \equiv sq$, $P \equiv wq$. Then (35) is the product of $M - tq \not\equiv 0$ by $sV - wR \equiv 0$. By using the conditions which follow readily from $L \equiv sq$, $P \equiv wq$, and from the terms free of t in $sV - wR$, we find that the latter is the product of q by a determinant which is evidently zero. Hence the left rank is 3 for the algebra

$$\begin{aligned} &e_1^2 = e_2, \quad e_1e_2 = a + be_1, \quad e_2e_1 = b_0 + b_1e_1 + b_2e_2, \quad e_2^2 = c_0 + c_2e_2, \\ (36)\quad &e_2e_3 = d_0 + d_2e_2 + b_1e_3, \quad e_1e_3 = a_0 + a_3e_3, \quad e_3e_2 = g_0 + g_2e_2 + be_3, \\ &e_3e_1 = f_0 + f_1e_1 + f_2e_2 - a_3e_3, \quad e_3^2 = h_0 + h_2e_2 + f_1e_3. \end{aligned}$$

Then X^2, e_2X and Xe_2 are linear functions of 1, e_2, X. Hence if Y is any element, 1, e_2, Y give a basis of a sub-algebra. Hence algebra (36) has left rank 3 and right rank 3. For, from the linear function of 1, e_2, Y giving Y^2Y (or YY^2) we may eliminate e_2 by means of $Y^2 = g + he_2 + jY$, $h \neq 0$.

13. Conversely, for every element Y of an algebra A of order 4 over a field F, let 1, E, Y give a basis of a sub-algebra, where E is fixed and is not a multiple of 1. The right and left ranks of A are $\leqq 3$. Exclude A's of rank 2. We choose an element e not satisfying a quadratic equation in A. The case $Y = E$ shows that E satisfies a quadratic equation in A. Hence 1, e, E are linearly independent in F. The case $Y = e$ gives $e^2 = r + se + tE$, $t \neq 0$. Write

$$e_1 = e - s/2, \qquad e_2 = e_1^2 = r + s^2/4 + tE.$$

Then 1, e_2, Y give a basis of our sub-algebra. Take (32) as a new e_3. We find that 1, e_2, Y form an algebra for every Y in A if and only if A is identical with algebra (36) with the second equation replaced by $e_1e_2 = a + be_1 + ce_2$.

14. Modify §12 by taking $L \equiv sQ$, $P \equiv st + wQ$. Then (35) holds (whence the left rank is 3) if and only if algebra (33) becomes

$$\begin{aligned} &e_1^2 = e_2, \quad e_1e_2 = a + be_1, \quad e_1e_3 = a_0, \quad e_2e_1 = b_0 + b_1e_1 + b_2e_2 + e_3, \\ (37) \quad &e_2^2 = c_0 + (b + b_1)e_2, \quad e_2e_3 = d_0 + b_1e_3, \quad e_3^2 = h_0 + f_1e_3, \\ &e_3e_2 = g_0 + f_1e_2 + be_3, \quad e_3e_1 = f_0 + f_1e_1 + f_2e_2. \end{aligned}$$

The right rank is 4, as shown by the right minimum equation for e_1. Note the sub-algebra (1, e_2, e_3).

Part IV. Division algebras

15. Let J be a root of a cyclic equation

$$x^n - c_1x^{n-1} + \cdots + (-1)^n c_n = 0,$$

i.e., an equation irreducible in a field F having the roots

$$J,\ J' = \theta(J),\ J'' = \theta(J') = \theta^2(J),\ \cdots,\ J^{(n-1)} = \theta^{n-1}(J),\ [\theta^n(J) = J],$$

where superscripts on θ denote iteratives (and not powers), while θ is a polynomial with coefficients in F. Write $B' = B(J)$, $B'' = B(J')$, etc. Define an algebra with the elements $A + BI$, where A and B are polynomials in J with coefficients in F, with the law of multiplication

$$(38) \quad (A + BI)(X + YI) = R + SI, \quad R = AX + gBY'J, \quad S = BX' + AY,$$

where g is a number $\neq 0$ of F. Both distributive laws hold.

Given R, S and X, Y not both zero, we can find A, B uniquely if

$$(39) \qquad \begin{vmatrix} X & gY'J \\ Y & X' \end{vmatrix} \neq 0.$$

This holds if $Y = 0$. Next, let $Y \neq 0$, $XX' = gYY'J$. This fails if $n = 2$

since J is not in F. If $n \geqq 3$, we take the product of its conjugates, write $N(X)$ for the norm $XX' \cdots X^{(n-1)}$ of X, and get

$$N^2(X) = g^n c_n N^2(Y),$$

which fails if $g^n c_n$ is not the square of the norm of a number of $F(J)$.

Given R, S and A, B not both zero, we seek X, Y. But

$$C \equiv gBS'J - A'R = gBB'JX'' - AA'X. \tag{40}$$

If $n = 2$, then $X'' = X$ and we obtain X uniquely and then Y. If $n \geqq 3$, let D denote the determinant of the coefficients of $X, X', \cdots, X^{(n-1)}$ in $C, C', \cdots, C^{(n-1)}$.

If n is odd, we find that

$$D = -N^2(A) + g^n c_n N^2(B). \tag{41}$$

But if n is even, $n \geqq 4$, we find that $D = UU'$, where

$$U = -N(A) + g^{n/2}N(B)JJ'' \cdots J^{(n-2)}. \tag{42}$$

Transpose and multiply by the conjugate equation.

THEOREM 3. *Let $(-1)^n c_n$ be the constant term of a cyclic equation of degree n over a field F. The algebra in which multiplication is defined by* (38) *is a division algebra if $n = 2$, and when $n \geqq 3$ if $g^n c_n$ is not the square of the norm of a number of the field $F(J)$.*

For $n = 2$, the algebra over F has the basis $1, J, I, K$, where

$$\begin{aligned} &J^2 = cJ - d, \quad I^2 = gJ, \quad JI = K, \quad K^2 = dgJ, \quad IK = gd, \quad KJ = dI, \\ &IJ = cI - K, \quad JK = cK - dI, \quad KI = gcJ - gd. \end{aligned} \tag{43}$$

This is a division algebra if $x^2 - cx + d = 0$ is irreducible in F; it is remarkable that there is no further condition.

A generalization of (38) is

$$\begin{aligned} R &= AX + (aBY + bBY' + cB'Y + dB'Y')J, \\ S &= AY + mBX + nBX', \qquad m + n = 1. \end{aligned} \tag{44}$$

If $a = b = c = n = 0$, $d = m = 1$, we have the commutative algebra cited in §1, which is a division algebra if c_n is not the square of the norm of a number of $F(J)$.

16. Algebras of order 2 without a modulus.

I. Let e^2 be a scalar multiple of e for every element e. As in the proof of Lemma 1,

$$e_1^2 = ae_1, \qquad e_2^2 = be_2, \qquad e_1e_2 + e_2e_1 = be_1 + ae_2.$$

Hence if $e = xe_1 + ye_2$, $e^2 = (ax + by)e$.

Let a and b be not both zero. By symmetry, take $a \neq 0$. Taking e_1/a as a new e_1, we have $a = 1$. The idempotent elements

$$e = (1 - by)e_1 + ye_2, \qquad E = (1 - bY)e_1 + Ye_2$$

are independent if $Y \neq y$. Let $e_1e_2 = je_1 + ke_2$. The conditions for $eE = 0$ are

$$1 - bY + Yj - yj = 0, \qquad Yk + y - yk = 0.$$

These have unique solutions y, Y with $Y \neq y$ if $j + bk - b \neq 0$. Then

$$e^2 = e, \qquad E^2 = E, \qquad eE = 0, \qquad Ee = e + E. \tag{45}$$

II. In the contrary case, we have (20)–(26) after suppressing scalars and the first equation (23). The condition for $S = J$ is now a homogeneous quadratic in b, c. Instead, we attempt to make $T = K$, given $t \neq k$; the condition is $b = mc$, $m = (s - j)/(t - k)$. Then $D \neq 0$ if $c \neq 0$ and

$$m^3 + m^2(k + t) + m(p - j - s) - n \neq 0. \tag{46}$$

In this case we may therefore take $t = k$.

UNIVERSITY OF CHICAGO.

AN APPLICATION OF LAGUERRE POLYNOMIALS

By D. V. Widder

Introduction. By the Laguerre polynomial of order n we mean the polynomial of degree n

$$L_n(x) = \sum_{k=0}^{n} \binom{n}{k} \frac{(-x)^k}{k!}. \tag{1}$$

These polynomials are known to form an orthogonal set with respect to the weight function e^{-x} on the interval $(0, \infty)$, that is,

$$\int_0^\infty e^{-x} L_m(x)\, L_n(x)\, dx = \begin{matrix} 0 \\ 1 \end{matrix} \qquad \begin{matrix} (m \neq n) \\ (m = n). \end{matrix}$$

The very form of this integral suggests that there should be an intimate relation between Laguerre polynomials and Laplace integrals

$$f(x) = \int_0^\infty e^{-xt}\, d\alpha(t). \tag{2}$$

It is the purpose of the present paper to examine more closely this relationship.

We are able by use of the Laguerre polynomials to give a new proof of a theorem of S. Bernstein to the effect that every completely monotonic function $f(x)$,

$$(-1)^k f^{(k)}(x) \geqq 0 \qquad (x > 0, k = 0, 1, 2, \cdots),$$

can be represented in the form (2), where $\alpha(t)$ is a non-decreasing function. The vital part of the proof is based on the known result that the function

$$K(x, y, t) = \sum_{n=0}^{\infty} L_n(x)\, L_n(y)\, t^n$$

is non-negative for $0 \leqq x < \infty$, $0 \leqq y < \infty$, $0 \leqq t < 1$. This fact was proved by G. H. Hardy in 1932 and by G. N. Watson in 1933. Incidentally we obtain a new inversion formula for (2) by a series of Laguerre polynomials. Our method shows automatically that the series in question is summable in the sense of Abel to the function $\alpha(t)$. It can be seen by other considerations that the series actually converges to $\alpha(t)$.

1. **The operator I.** We begin by defining an additive operator I which operates for the present only on functions of the form $e^{-x} P(x)$ when $P(x)$ is a

Received March 2, 1935; presented to the American Mathematical Society April 20, 1935.

polynomial. Let $f(x)$ be completely monotonic in the interval $0 \leqq x < \infty$, that is,

$$(-1)^k f^{(k)}(x) \geqq 0 \qquad (x > 0;\ k = 0, 1, 2, \cdots),$$

$$f(0) = \lim_{x \to 0+} f(x) = A,$$

where A is some finite constant, which will clearly be non-negative. Then we make the

DEFINITION. *If $P(x)$ is the polynomial*

$$P(x) = \sum_{k=0}^{n} a_k x^k,$$

then

$$I[e^{-x}P(x)] = \sum_{k=0}^{n} a_k(-1)^k f^{(k)}(1).$$

It is immediately apparent that if $P_1(x)$ and $P_2(x)$ are arbitrary polynomials and c_1 and c_2 are arbitrary constants

$$I[c_1e^{-x}P_1(x) + c_2e^{-x}P_2(x)] = c_1I[e^{-x}P_1(x)] + c_2I[e^{-x}P_2(x)].$$

Furthermore, we can show I to be a positive operator in the sense of

THEOREM 1. *If $P(x)$ is a polynomial such that*

$$P(x) \geqq 0 \qquad (0 \leqq x < \infty), \tag{1.1}$$

then

$$I[e^{-x}P(x)] \geqq 0.$$

To prove this result we first recall[1] that any polynomial satisfying (1.1) can be written in the form

$$P(x) = [A(x)]^2 + [B(x)]^2 + x[D(x)]^2 + x[E(x)]^2,$$

where $A(x)$, $B(x)$, $D(x)$, and $E(x)$ are suitably chosen polynomials. Hence, it will be sufficient to prove for an arbitrary polynomial $P(x)$ that

$$I[e^{-x}P^2(x)] \geqq 0,$$

$$I[e^{-x}xP^2(x)] \geqq 0.$$

Since $f(x)$ is completely monotonic, it is easily seen[2] that the quadratic forms

[1] See, for example, G. Pólya and G. Szegö, *Aufgaben und Lehrsätze aus der Analysis*, vol. II, p. 82, No. 45.

[2] See D. V. Widder, *Necessary and sufficient conditions for the representation of a function by a Laplace integral*, Transactions of the American Mathematical Society, vol. 33 (1931), p. 855. It should be understood that the non-negative character of the determinants there considered is not sufficient to insure that the forms (1.2) are non-negative. However, if the same limit process employed in the article cited is applied directly to quadratic forms, the result (1.2) is obtained.

$$(1.2)\qquad \begin{array}{c}\sum_{j=0}^{n}\sum_{k=0}^{n}(-1)^{j+k}f^{(j+k)}(1)\xi_j\xi_k \\ \sum_{j=0}^{n}\sum_{k=0}^{n}(-1)^{j+k+1}f^{(j+k+1)}(1)\xi_j\xi_k\end{array}\qquad (n = 0, 1, 2, \cdots)$$

are non-negative for all values of the variables ξ_k.

Hence, if

$$P(x) = \sum_{k=0}^{n} a_k x^k,$$

then

$$I[e^{-x}P^2(x)] = \sum_{j=0}^{n}\sum_{k=0}^{n} a_j a_k(-1)^{j+k} f^{(j+k)}(1) \geqq 0,$$

$$I[e^{-x}x\,P^2(x)] = \sum_{j=0}^{n}\sum_{k=0}^{n} a_j a_k(-1)^{j+k+1} f^{(j+k+1)}(1) \geqq 0.$$

This completes the proof of the theorem.

2. **Extension of the domain of the operator** I. In order to extend slightly the domain of functions to which the operator I is applicable we need to prove

THEOREM 2. *If $P(x)$ is a polynomial such that*

$$(2.1)\qquad P(x) < e^x \qquad (0 \leqq x < \infty),$$

then it is possible to determine an integer m such that

$$P(x) < \sum_{k=0}^{m}\frac{x^k}{k!} \qquad (0 \leqq x < \infty).$$

We begin the proof by defining the polynomials

$$Q_n(x) = \sum_{k=0}^{n}\frac{x^k}{k!}.$$

If n is the degree of $P(x)$, then

$$\lim_{x\to\infty}\frac{P(x)}{Q_{n+1}(x)} = 0.$$

Hence there exists a positive number R such that

$$P(x) < Q_{n+1}(x) \qquad (x \geqq R).$$

But

$$(2.2)\qquad \lim_{k\to\infty} Q_{n+k}(x) = e^x$$

uniformly in $0 \leqq x \leqq R$. By the hypothesis (2.1), there exists a positive number ϵ such that

$$P(x) < e^x - \epsilon \qquad (0 \leqq x \leqq R).$$

By (2.2) it is possible to determine, corresponding to this ϵ, a positive integer q such that

$$Q_{n+q}(x) > e^x - \epsilon \qquad (0 \leqq x \leqq R).$$

Hence

$$P(x) < Q_{n+q}(x) \qquad (0 \leqq x \leqq R).$$

But

$$Q_{n+1}(x) \leqq Q_{n+q}(x) \qquad (q \geqq 1, x \geqq R),$$

so that

$$P(x) < Q_m(x) \qquad (0 \leqq x < \infty)$$

if $m = n + q$, and our theorem is proved.

We now extend the definition of I so that it will apply in particular to constants and will retain its additive property.

Definition. *If c is any constant, then*

$$I[c + e^{-x} P(x)] = cf(0) + I[e^{-x} P(x)].$$

We show next that the operator remains positive after this extension.

Theorem 3. *If $P(x)$ is a polynomial such that*

$$c + e^{-x} P(x) \geqq 0 \qquad (0 \leqq x < \infty),$$

then

$$I[c + e^{-x} P(x)] \geqq 0.$$

We note first that $c \geqq 0$ since

$$\lim_{x \to \infty} e^{-x} P(x) = 0 .$$

Let η be an arbitrary positive constant. Then

$$\left.\begin{aligned} &\eta + c + e^{-x} P(x) > 0 , \\ &-\frac{P(x)}{\eta + c} < e^x \end{aligned}\right\} \qquad (0 \leqq x < \infty).$$

By Theorem 2 an integer m exists such that

$$-\frac{P(x)}{\eta + c} < \sum_{k=0}^{m} \frac{x^k}{k!} \qquad (0 \leqq x < \infty),$$

$$e^{-x}\left[\sum_{k=0}^{m} \frac{x^k}{k!} + \frac{P(x)}{\eta + c}\right] > 0 .$$

Hence, by Theorem 1

$$I[e^{-x}P(x)] + (\eta + c)\sum_{k=0}^{m}(-1)^k\frac{f^{(k)}(1)}{k!} \geqq 0,$$

$$-I[e^{-x}P(x)] \leqq (\eta + c)\sum_{k=0}^{\infty}(-1)^k\frac{f^{(k)}(1)}{k!},$$

provided that the infinite series on the right converges. That it does and has the value $f(0)$ follows[3] from the fact that $f(x)$ is completely monotonic for $x \geqq 0$. Hence we have

$$I[e^{-x}P(x)] + (\eta + c)f(0) \geqq 0,$$

or, since η was arbitrary,

$$I[e^{-x}P(x)] + cf(0) = I[c + e^{-x}P(x)] \geqq 0.$$

This is the desired result. From it follows at once the

COROLLARY. *The inequality*

$$|e^{-x}P(x)| \leqq M \qquad (0 \leqq x < \infty)$$

implies

$$|I[e^{-x}P(x)]| \leqq Mf(0).$$

For the hypothesis implies that

$$M + e^{-x}P(x) \geqq 0,$$

$$M - e^{-x}P(x) \geqq 0,$$

so that Theorem 3 is applicable.

3. **Application of I to the generating function of the Laguerre polynomials.** Results which we shall need regarding the generating function $K(x, y, t)$ of the Laguerre polynomials we summarize in

THEOREM 4. *The series*

$$e^{-x}K(x,y,t) = \sum_{n=0}^{\infty} e^{-x}L_n(x)L_n(y)t^n \tag{3.1}$$

converges uniformly to a non-negative sum in the interval $0 \leqq x < \infty$ for each positive y and each positive t less than unity. Moreover, the sequence of constants

$$\lambda_n = I[e^{-x}L_n(x)] = \sum_{k=0}^{n}\binom{n}{k}\frac{f^{(k)}(1)}{k!} \qquad (n = 0, 1, 2, \cdots)$$

[3] Serge Bernstein, *Sur la définition et les propriétés des fonctions analytiques d'une variable réelle*, Mathematische Annalen, vol. 75 (1914), p. 449. The analyticity of $f(x)$ insures that the Taylor development for $f(x)$ at the point $x = 1$ converges to $f(x)$ for $|x - 1| < 1$. That it also converges for $x = 0$ follows by a Tauberian theorem since the coefficients are positive and since $f(0+) = A$.

satisfies the relation

$$|\lambda_n| \leqq f(0) \qquad (n = 0, 1, 2, \cdots). \tag{3.2}$$

To show that the series (3.1) converges uniformly in λ for specified values of y and t we make use of the known[4] inequality

$$| e^{-x/2} L_n(x) | \leqq 1 \qquad (x \geqq 0), \tag{3.3}$$

and thus obtain

$$\sum_{n=0}^{\infty} e^{-x} L_n(x) L_n(y_0) t_0^n \ll e^{y_0} \sum_{n=0}^{\infty} t_0^n \qquad (x \geqq 0, y_0 > 0, 0 < t_0 < 1).$$

The dominant series is independent of x and converges.

That

$$e^{-x} K(x, y, t) \geqq 0 \qquad (x > 0, y > 0, 0 < t < 1)$$

follows from a result of G. H. Hardy.[5]

Finally, to prove (3.2) we have only to make use of (3.3) and apply the Corollary of Theorem 3:

$$| \lambda_n | = | I[e^{-x} L_n(x)] | \leqq f(0).$$

We now apply the operator I to the function $e^{-x} K(x, y, t)$ considered as a function of x. Since $K(x, y, t)$ is not a polynomial, the operation is not actually defined. However, without extending the field of the operation we may consider the function

$$A_t(y) = \sum_{n=0}^{\infty} \lambda_n L_n(y) t^n$$

on its own merits. We should expect that it would be a non-negative function, and this we prove in

THEOREM 5. *The series*

$$A_t(y) = \sum_{n=0}^{\infty} \lambda_n L_n(y) t^n$$

converges to a non-negative value for $0 \leqq y < \infty$, $0 \leqq t < 1$. *Moreover,*

$$| A_t(y) | \leqq \frac{f(0)\, e^{y/2}}{1 - t} \qquad (0 \leqq y < \infty, 0 \leqq t < 1). \tag{3.4}$$

[4] G. Szegö, *Ein Beitrag zur Theorie der Polynome von Laguerre und Jacobi*, Mathematische Zeitschrift, vol. 1 (1918), p. 341.

[5] G. H. Hardy, *Summation of a series of polynomials of Laguerre*, Journal of the London Mathematical Society, vol. 7 (1932), p. 138. See also G. N. Watson, *Notes on generating functions of polynomials: (1) Laguerre polynomials*, Journal of London Mathematical Society, vol. 8 (1933), p. 189.

For because of Theorem 4 and (3.3)

$$|A_t(y)| \leqq \sum_{n=0}^{\infty} f(0)e^{y/2}t^n = \frac{f(0)e^{y/2}}{1-t} \qquad (0 \leqq y < \infty, 0 \leqq t < 1).$$

Since series (3.1) converges uniformly in x, then corresponding to an arbitrary positive number η there is an integer m_0 independent of x such that

$$\eta + \sum_{n=0}^{m} L_n(x)e^{-x}L_n(y_0)t_0^k > 0 \qquad (0 \leqq x < \infty, y_0 > 0, 0 < t_0 < \infty).$$

Hence, by Theorem 3

$$\eta f(0) + \sum_{n=0}^{m} \lambda_n L_n(y_0)t_0^k \geqq 0,$$

$$\eta f(0) + \sum_{n=0}^{\infty} \lambda_n L_n(y_0)t_0^k \geqq 0.$$

Since η is arbitrary,

$$A_{t_0}(y_0) \geqq 0,$$

so that the theorem is established.

4. **Bernstein's Theorem.** We are now able to prove

THEOREM 6. *The integral*

$$\int_0^{\infty} e^{-xy}A_t(y)\,dy$$

converges for $x > \frac{1}{2}$ *and*

$$f(x) = \lim_{t\to 1-}\int_0^{\infty} e^{-xy}A_t(y)\,dy \qquad (1 < x < 2).$$

The convergence of the integral for $x > \frac{1}{2}$ follows at once from (3.4). To find its value we have

$$(4.1)\quad \int_0^{\infty} e^{-xy}A_t(y)\,dy = \int_0^{\infty} e^{-xy}\sum_{n=0}^{\infty}\lambda_n L_n(y)t^n\,dy = \sum_{n=0}^{\infty}\lambda_n t^n \frac{(x-1)^n}{x^{n+1}} \qquad (\tfrac{1}{2} < x),$$

for

$$\int_0^{\infty} e^{-xy}L_n(y)\,dy = \frac{(x-1)^n}{x^{n+1}}.$$

To justify the above term-by-term integration we have only to employ (3.4) and note that the integral

$$\int_0^{\infty} e^{-xy}\frac{f(0)e^{y/2}}{1-t}\,dy$$

converges for $x > \frac{1}{2}$.

Now the series (4.1) can be transformed as follows:

$$
\begin{aligned}
\sum_{n=0}^{\infty} \lambda_n t^n \frac{(x-1)^n}{x^{n+1}} &= \frac{1}{x}\sum_{n=0}^{\infty}\left(t-\frac{t}{x}\right)^n \sum_{k=0}^{n}\binom{n}{k}\frac{f^{(k)}(1)}{k!} \\
&= \frac{1}{x}\sum_{k=0}^{\infty}\frac{f^{(k)}(1)}{k!}\sum_{n=k}^{\infty}\binom{n}{k}\left(t-\frac{t}{x}\right)^n \qquad (4.2) \\
&= \sum_{k=0}^{\infty}\frac{f^{(k)}(1)}{k!}\frac{(tx-t)^k}{(x+t-tx)^{k+1}}.
\end{aligned}
$$

The interchange in the order of summation is permissible since the double series converges absolutely when $1 \leqq x < 2$. For

$$\sum_{n=0}^{\infty}\left(t-\frac{t}{x}\right)^n \sum_{k=0}^{n}\binom{n}{k}\frac{f^{(k)}(1)}{k!} \ll \sum_{n=0}^{\infty}\left(t-\frac{t}{x}\right)^n 2^n M,$$

since the convergence of Taylor's series for $f(x)$ about $x = 1$ at the point $x = 0$ implies the existence of a constant M such that

$$(-1)^k f^{(k)}(1) < Mk! \qquad (k = 0, 1, 2, \cdots).$$

The dominant series clearly converges since

$$0 \leqq \left(t-\frac{t}{x}\right)^n 2^n < 1 \qquad (0 < t < 1, 1 \leqq x < 2).$$

Finally, by letting t approach unity in (4.2) we obtain

$$\lim_{t\to 1-}\int_0^{\infty} e^{-xt} A_t(y)\,dy = \sum_{k=0}^{\infty}\frac{f^{(k)}(1)}{k!}(x-1)^k = f(x).$$

To justify this limit process we note that the relation

$$\sum_{k=0}^{\infty}\frac{f^{(k)}(1)}{k!}\left[\frac{t(x-1)}{x+t-tx}\right]^k \ll \sum_{k=0}^{\infty}(-1)^k f^{(k)}(1)\frac{(x-1)^k}{k!} \qquad (1 \leqq x < 2, 0 \leqq t \leqq 1)$$

shows that the series (4.2) converges uniformly for $0 \leqq t \leqq 1$. This completes the proof. We now establish

THEOREM 7. *If $f(x)$ is completely monotonic for $0 \leqq x < \infty$, then*

$$f(x) = \int_0^{\infty} e^{-xt}\,d\alpha(t) \qquad (x \geqq 0),$$

where $\alpha(t)$ is a uniformly bounded non-decreasing function for $0 \leqq t < \infty$.

For, set

$$\alpha_t(x) = \int_0^{x} e^{-y} A_t(y)\,dy.$$

Since $A_t(y)$ is non-negative, $\alpha_t(y)$ is non-decreasing. Moreover, (4.1) shows that

$$0 \leqq \alpha_t(x) \leqq \lambda_0 = f(0) \qquad (0 \leqq t < 1, 0 \leqq x < \infty).$$

Hence by a result of E. Helly[6] we may pick a sequence of numbers $0 < t_1 < t_2 < \cdots < 1$ such that

$$\lim_{j\to\infty} t_j = 1,$$

$$\lim_{j\to\infty} \alpha_{t_j}(x) = \bar{\alpha}(x) \qquad (0 \leqq x < \infty),$$

where $\bar{\alpha}(x)$ is a suitable non-decreasing uniformly bounded function. Hence, by Theorem 6 we have for $1 < x < 2$

$$f(x) = \lim_{j\to\infty} \int_0^\infty e^{-(x-1)y}\, d\alpha_{t_j}(y) = \lim_{t\to 1-} \int_0^\infty e^{-xy} A_t(y)\, dy.$$

Finally, by the Helly-Bray theorem,[7] we may take the limit under the integral sign and obtain

$$\text{(4.3)} \qquad \begin{aligned} f(x) &= \int_0^\infty e^{-(x-1)y}\, d\bar{\alpha}(y) \qquad (1 < x < 2) \\ &= \int_0^\infty e^{-xy}\, d\alpha(y) \end{aligned}$$

where

$$\alpha(y) = \int_0^y e^x\, d\bar{\alpha}(x) \qquad (y > 0),$$

$$\alpha(0) = 0.$$

But the integral (4.3) represents an analytic function at least for $x > 1$ which must consequently coincide with the analytic function $f(x)$ there. Since $\alpha(y)$ is non-decreasing, the function represented by (4.3) must have a singularity at the real point on its axis of convergence.[8] Since $f(x)$ has no singularity for $x > 0$, we see that the integral (4.3) must converge to $f(x)$ for $x \geqq 0$ and our theorem is proved. It is to be noted that $\alpha(y)$ is bounded since $f(0) = \alpha(\infty)$.

5. **Remarks.** The case in which $f(x)$ is completely monotonic in the open interval $0 < x < \infty$ may be obtained from Theorem 7 by applying it to the

[6] E. Helly, *Über lineare Funktionaloperationen*, Wiener Sitzungsberichte, vol. 121 (1921), p. 265.

[7] See, for example, G. C. Evans, *The Logarithmic Potential, Discontinuous Dirichlet and Neumann Problems*, Colloquium Publications, vol. 6, of the American Mathematical Society, 1927, p. 15.

[8] D. V. Widder, *A generalization of Dirichlet's series and of Laplace's integrals by means of a Stieltjes integral*, Transactions of the American Mathematical Society, vol. 31 (1929), p. 719.

function $f(x+\epsilon)$ and by using the fact that the representation of a function $f(x)$ by a Laplace integral is essentially unique. The function $\alpha(t)$ corresponding to $f(x)$ will in general be unbounded.

We also remark that it can be shown that

$$\lim_{t\to 1-} \alpha_t(x) = \bar{\alpha}(x)$$

at least for all points of continuity of $\bar{\alpha}(x)$. The argument is similar to one employed by F. Hausdorff[9] and is omitted. Hence

$$\alpha(y) = \int_0^y e^u d\bar{\alpha}(u) = \lim_{t\to 1-} \int_0^y e^u d\alpha_t(u) = \lim_{t\to 1-} \int_0^y A_t(u)du .$$

But

$$\int_0^x A_t(y)dy = \sum_{n=0}^{\infty} \lambda_n t^n \int_0^x L_n(y)dy ,$$

so that the series

$$\sum_{n=0}^{\infty} \lambda_n \int_0^x L_n(y)dy \tag{5.1}$$

is summable in the sense of Abel to $\alpha(x)$ at points of continuity of $\alpha(x)$.

By an appeal to the general theory of Laguerre polynomials one may show that (5.1) actually converges for all finite $x \geqq 0$ to the normalized function $\alpha^*(x)$,

$$\alpha^*(0) = 0 , \; \alpha^*(x) = \frac{\alpha(x+) + \alpha(x-)}{2} \qquad (x > 0) .$$

Indeed, we have[10]

$$\int_0^x L_n(y)dy = L_n(x) - L_{n+1}(x) .$$

By partial summation

$$\sum_{n=0}^{m} \lambda_n[L_n(x) - L_{n+1}(x)] = \lambda_0 L_0(x) + \sum_{n=1}^{m} [\lambda_n - \lambda_{n-1}]\, L_n(x) = - \lambda_m L_{m+1}(x) .$$

Since

$$\lambda_m = O(1) \qquad (m \to \infty) ,$$

$$L_{m+1}(x) = O(m^{-1/4}) \qquad (m \to \infty)$$

for each fixed positive x, we have

$$\lim_{m\to\infty} \lambda_m L_{m+1}(x) = 0 ,$$

[9] F. Hausdorff, *Momentprobleme für ein endliches Intervall*, Mathematische Zeitschrift, vol. 16 (1923), p. 226.

[10] For the formulas and results stated in the rest of this paper see G. Szegö, *Beiträge zur Theorie der Laguerreschen Polynome*, Mathematische Zeitschrift, vol. 25 (1926), p. 87.

so that it will be sufficient to discuss the convergence behavior of the series

$$\lambda_0 L_0(x) + \sum_{n=1}^{\infty} [\lambda_n - \lambda_{n-1}]\, L_n(x)\,. \tag{5.2}$$

But

$$\lambda_n = \int_0^\infty e^{-y}\, L_n(y) d\alpha(y) = \int_0^\infty e^{-y}\, [L_n(y) - L_n'(y)]\, \alpha(y) dy\,,$$

$$\lambda_n - \lambda_{n-1} = \int_0^\infty e^{-y}\, [L_n(y) - L_{n-1}(y) - L_n'(y) + L_{n-1}'(y)]\, \alpha(y) dy\,,$$

$$= \int_0^\infty e^{-y}\, L_n(y)\, \alpha(y) dy \qquad (n = 1, 2, \cdots)\,,$$

$$\lambda_0 = \int_0^\infty e^{-y}\, L_0(y)\, \alpha(y) dy\,.$$

Hence (5.2) becomes

$$\sum_{n=0}^{\infty} L_n(x) \int_0^\infty e^{-y}\, \alpha(y)\, L_n(y) dy\,,$$

a series which is known to converge to the uniformly bounded non-decreasing function $\alpha^*(x)$ for all $x \geqq 0$.

We thus have an inversion of the Laplace integral under the present hypotheses. That is

$$\alpha(x) = \sum_{n=}^{\infty} \int_0^x L_n(y) dy \sum_{k=0}^{n} \binom{n}{k} \frac{f^{(k)}(1)}{k!}\,.$$

It is of interest that the formula involves a knowledge of the function $f(x)$ only in a neighborhood of $x = 1$, whereas the inversion formula formerly studied by the author[11] involved a knowledge of the function in a neighborhood of $x = +\infty$.

We note that the operator I employed in the present paper can now be identified with the operation of integration with respect to $\alpha(t)$ from zero to infinity:

$$I[e^{-x}\, P(x)] = \int_0^\infty e^{-x}\, P(x) d\alpha(x)\,.$$

Our final remark is that one might use the more general Laguerre polynomials $L_n^{(\alpha)}(x)$ in the above development.

Harvard University.

[11] D. V. Widder, *The inversion of the Laplace integral and the related moment problem*, Transactions of the American Mathematical Society, vol. 36 (1934), p. 107.

ON CERTAIN FUNCTIONS CONNECTED WITH POLYNOMIALS IN A GALOIS FIELD

By Leonard Carlitz

1. Introduction. Let $GF(p^n)$ denote a fixed Galois field[1] of order p^n; let $E = E(x)$ denote a polynomial in an indeterminate x with coefficients in $GF(p^n)$. Consider the product $\psi_k(t) = \Pi(t - E)$, extended over all E of degree $< k$, where k is an arbitrary positive integer. We show, to begin with, that the product has the expansion

$$\sum_{j=0}^{k} (-1)^j \begin{bmatrix} k \\ j \end{bmatrix} t^{p^{nj}}, \tag{1.01}$$

the coefficients (defined explicitly in §2) having certain properties analogous to those of the binomial coefficients. Of the properties of $\psi_k(t)$, it is evident from the form of (1.01) that, for c in $GF(p^n)$,

$$\psi_k(ct) = c\psi_k(t), \qquad \psi_k(t + u) = \psi_k(t) + \psi_k(u);$$

we accordingly call $\psi_k(t)$ a *linear* polynomial.[2] As a second characteristic property we mention

$$\psi_k(xt) - x\psi_k(t) = (x^{p^{nk}} - x)\psi_{k-1}^{p^n}(t).$$

This relation suggests the study of the operator Δ defined by

$$\Delta f(t) = f(xt) - xf(t), \tag{1.02}$$

where $f(t)$ is a linear polynomial. See §3.

We suppose next that k in (1.01) becomes infinite; the product $\Pi(t - E)$ must be modified somewhat. Actually we consider

$$\Pi'\left(1 - \frac{t^{p^n-1}}{E^{p^n-1}}\right), \tag{1.03}$$

the product extending over all *primary* E, that is, over all polynomials in which the coefficient of the highest power of x is the 1 element of the Galois field. As we shall see, the question of convergence causes little difficulty; we find that the infinite product (1.03) has the expansion

$$\frac{1}{\xi}\psi(t\xi) = \frac{1}{\xi}\sum_{k=0}^{\infty} \frac{(-1)^k}{F_k} t^{p^{nk}} \xi^{p^{nk}}, \tag{1.04}$$

Received April 17, 1935.

[1] For the properties of Galois fields assumed here, see L. E. Dickson, *Linear Groups*, 1901, pp. 3–54.

[2] Called p-polynomials by O. Ore, Transactions of the American Mathematical Society, vol. 35 (1933), pp. 559–584.

where

$$F_k = [k][k-1]^{p^n} \cdots [1]^{p^{n(k-1)}}, \qquad [k] = x^{p^{nk}} - x,$$

and

$$\xi = \lim_{k\to\infty} \frac{[1]^{p^{nk}/(p^n-1)}}{[k][k-1]\cdots[1]}.$$

The function $\psi(t)$ has, first, the linearity properties

$$\psi(ct) = c\psi(t), \qquad \psi(t+u) = \psi(t) + \psi(u).$$

In addition,

$$\Delta\psi(t) = \psi(xt) - x\psi(t) = -\psi^{p^n}(t),$$

and

$$\psi(t + E\xi) = \psi(t), \tag{1.05}$$

for all polynomials E. According to (1.05), $\psi(t)$ has the period ξ (and it is not difficult to show that it has no other period). On the other hand, since the 'numerical' quantities appearing in the definition of $\psi(t)$ are all in $GF(p)$, we may describe $\psi(t)$ as an n-ply periodic function. This is clear if we put (1.05) in the form

$$\psi(t + E_1\xi + E_2\vartheta\xi + \cdots + E_n\vartheta^{n-1}\xi) = \psi(t),$$

where the coefficients of the E_i are in $GF(p)$, and ϑ defines the $GF(p^n)$. As we shall show elsewhere, $\psi(t)$ is but the simplest of an extensive class of n-ply periodic linear functions.

If $f(t)$ and $g(t)$ are linear functions such that $f(g(t)) \equiv t$, then g is the inverse of f. It is easily shown that in this event f is also the inverse of g; as we shall see, the matter is independent of the question of convergence. For the function $\psi(t)$ we find the inverse $\lambda(t)$, defined by

$$\lambda(t) = \sum_{k=0}^{\infty} \frac{t^{p^{nk}}}{[k][k-1]\cdots[1]}. \tag{1.06}$$

Explicit formulas are also obtained for the inverse of $\psi_k(t)$.

We next study the reciprocals of ψ_k and ψ. This leads to the introduction of certain rational fractions in x analogous to the Bernoulli numbers, which in turn enable us to evaluate

$$\sum_{\deg E = k}{}' E^{-(p^n-1)m} \qquad \text{and} \qquad \sum_{E}{}' E^{-(p^n-1)m},$$

the summations extending respectively over primary E of degree k, and over *all* primary polynomials E. In particular, for $m = (p^{nk}-1)/(p^n-1)$, we find that

$$\sum_{E}{}' E^{-(p^{nk}-1)} = \frac{\xi^{p^{nk}-1}}{[k][k-1]\cdots[1]}.$$

In the remainder of the paper we apply some of our earlier formulas to two types of higher congruences. The first type (§10) is the binomial; we give a new proof of F. K. Schmidt's theorem of higher reciprocity. The second type (§11) is the congruence

$$t^{p^n} - t \equiv A \pmod{P},$$

where P is irreducible of degree m. Our main theorem is *that the congruence is solvable if and only if AP' is congruent* (mod P) *to a polynomial of degree* $< m - 1$; here P' is the derivative of P with respect to x.

2. **The polynomial $\psi_k(t)$.** It will be convenient to introduce certain notations. We shall use c, c_j to denote elements of the $GF(p^n)$. $E, F, \cdots$ will denote polynomials:

$$E = c_0x^k + c_1x^{k-1} + \cdots + c_k;$$

for $c_0 = 1$, E is said to be *primary*; we write $k = \deg E$. We shall have frequent occasion to use sums and products taken over sets E satisfying certain conditions; we shall use the convention that Σ' and Π' indicate that only primary E be admitted. The polynomial 0 is not primary; ordinarily a sum or product over all polynomials of degree less than some k will include the polynomial 0.

Suppose now that $t_0, t_1, \cdots, t_k$ are $k + 1$ indeterminates, and that in the linear form

$$l = c_0t_0 + c_1t_1 + \cdots + c_kt_k,$$

the non-vanishing c_j of least subscript is the unit element of the Galois field. There are evidently

$$p^{nk} + p^{n(k-1)} + \cdots + 1 = \frac{p^{n(k+1)} - 1}{p^n - 1}$$

distinct linear forms l. Their product, by a formula due to E. H. Moore,[3] may be expressed as a determinant of order $k + 1$:

$$D(t_0, t_1, \cdots, t_k) = \left| t_i^{p^{n(k-j)}} \right| \qquad (i, j = 0, 1, \cdots, k). \tag{2.01}$$

If we restrict our attention to the linear forms in which $c_0 = 1$, then clearly

$$\prod_{(c)} (t_0 + c_1t_1 + \cdots + c_kt_k) = \frac{D(t_0, t_1, \cdots, t_k)}{D(t_1, t_2, \cdots, t_k)}, \tag{2.02}$$

the k-fold product extending over the p^{nk} sets $(c_1, c_2, \cdots, c_k)$ in the $GF(p^n)$. We now write t for t_0, and put

$$t_j = x^{k-j} \qquad (j = 1, \cdots, k),$$

[3] E. H. Moore, Bulletin of the American Mathematical Society, vol. 2 (1896), pp. 189–195.

then the identity (2.01) becomes

$$(2.03) \qquad \prod_{\deg E<k} (t+E) = \frac{D(t, x^{k-1}, \cdots, 1)}{D(x^{k-1}, \cdots, 1)},$$

the product extending over all polynomials E of degree $< k$. To evaluate the right member of (2.03), we first expand the determinant in the numerator with respect to the elements in the first row:

$$(2.04) \qquad D(t, x^{k-1}, \cdots, 1) = \sum_{j=0}^{k} (-1)^{k-j} t^{p^{nj}} D_j,$$

where D_j is the determinant of order k whose i^{th} row consists of

$$x^{(k-i)p^{nk}}, \cdots, x^{(k-i)p^{n(j+1)}}, x^{(k-i)p^{n(j-1)}}, \cdots, x^{k-i} \qquad (i = 1, \cdots, k).$$

The determinant D_j is evidently a Vandermonde determinant; it is therefore easily seen that

$$(2.05) \qquad \begin{aligned} D_j &= \prod_{h>i} (x^{p^{nh}} - x^{p^{ni}}) \qquad (h, i = 1, \cdots, k; h \neq j, i \neq j) \\ &= \frac{\prod_{h>i} (x^{p^{nh}} - x^{p^{ni}})}{\prod_{h>j} (x^{p^{nh}} - x^{p^{nj}}) \prod_{i<j} (x^{p^{nj}} - x^{p^{ni}})}, \end{aligned}$$

the products in the numerator extending over all $h, i = 1, \cdots, k$. If for brevity we put

$$(2.06) \qquad \begin{aligned} F_k &= (x^{p^{nk}} - x)(x^{p^{n(k-1)}} - x)^{p^n} \cdots (x^{p^n} - x)^{p^{n(k-1)}}, \\ L_k &= (x^{p^{nk}} - x)(x^{p^{n(k-1)}} - x) \cdots (x^{p^n} - x), \\ F_0 &= L_0 = 1, \end{aligned}$$

then (2.05) becomes

$$(2.07) \qquad D_j = \frac{F_k F_{k-1} \cdots F_1}{F_j L_{k-j}^{p^{nj}}}.$$

As for the denominator in the right member of (2.03), we find that

$$(2.08) \qquad D(x^{k-1}, \cdots, 1) = F_{k-1} F_{k-2} \cdots F_1.$$

Note that from the definition above, $D(x^{k-1}, \cdots, 1)$ is the product of all *primary* polynomials of degree $< k$; therefore F_k is the product of primary polynomials of degree $= k$. As for L_k, it may be shown[4] that it is the 'least common multiple' of polynomials of degree k.

Making use of formulas (2.04), (2.07) and (2.08), we see that the right member of (2.03) becomes

$$\sum_{i=0}^{k} (-1)^{k-i} \frac{F_k}{F_i L_{k-i}^{p^{ni}}} t^{p^{ni}}.$$

[4] Bulletin of the American Mathematical Society, vol. 38 (1932), pp. 736–744.

We shall find it convenient to use the symbol

$$\begin{bmatrix} k \\ i \end{bmatrix} = \frac{F_k}{F_i L_{k-i}^{p^{ni}}}, \qquad \begin{bmatrix} k \\ 0 \end{bmatrix} = \frac{F_k}{L_k}, \qquad \begin{bmatrix} k \\ k \end{bmatrix} = 1. \tag{2.09}$$

We may now state

THEOREM 2.1. *If E run through all the polynomials (including 0) of degree $< k$, then*

$$\psi_k(t) = \prod (t - E) = \sum_{i=0}^{k} (-1)^{k-i} \begin{bmatrix} k \\ i \end{bmatrix} t^{p^{ni}}, \tag{2.10}$$

where $\begin{bmatrix} k \\ i \end{bmatrix}$ *is defined by* (2.09). *$\psi_k(t)$ has the following properties:*

$$\psi_k(ct) = c\psi_k(t), \qquad c \text{ in } GF(p^n),$$

$$\psi_k(t + u) = \psi_k(t) + \psi_k(u),$$

$$\psi_k(E) = 0, \qquad \text{for deg } E < k.$$

It will be convenient to define $\psi_0(t) = t$.

We now derive certain properties of the coefficients $\begin{bmatrix} k \\ i \end{bmatrix}$. In the first place, by means of (2.06),

$$F_k = (x^{p^{nk}} - x)F_{k-1}^{p^n}, \qquad L_k = (x^{p^{nk}} - x)L_{k-1};$$

hence

$$\frac{F_k}{F_i L_{k-i}^{p^{ni}}} = \frac{x^{p^{nk}} - x}{x^{p^{ni}} - x}\left(\frac{F_{k-1}}{F_{i-1} L_{k-i}^{p^{n(i-1)}}}\right)^{p^n},$$

or more briefly

$$\begin{bmatrix} k \\ i \end{bmatrix} = \frac{x^{p^{nk}} - x}{x^{p^{ni}} - x} \begin{bmatrix} k-1 \\ i-1 \end{bmatrix}^{p^n}. \tag{2.11}$$

In the second place, from the obvious identity

$$x^{p^{nk}} - x = (x^{p^{ni}} - x) + (x^{p^{n(k-i)}} - x)^{p^{ni}},$$

formula (2.11) becomes

$$\begin{bmatrix} k \\ i \end{bmatrix} = \begin{bmatrix} k-1 \\ i-1 \end{bmatrix}^{p^n} + \frac{(x^{p^{n(k-i)}} - x)^{p^{ni}}}{x^{p^{ni}} - x} \frac{F_{k-1}^{p^n}}{F_{i-1}^{p^n} L_{k-i}^{p^{ni}}}$$

$$= \begin{bmatrix} k-1 \\ i-1 \end{bmatrix}^{p^n} + F_{k-1}^{p^n-1} \begin{bmatrix} k-1 \\ i \end{bmatrix}. \tag{2.12}$$

Making use of (2.11) and (2.12) we prove

THEOREM 2.2. *The polynomial $\psi_k(t)$ of Theorem 2.1 has the further properties:*

$$\psi_k(xt) - x\psi_k(t) = (x^{p^{nk}} - x)\psi_{k-1}^{p^n}(t), \tag{2.13}$$

$$\psi_k(t) = \psi_{k-1}^{p^n}(t) - F_{k-1}^{p^n-1} \cdot \psi_{k-1}(t). \tag{2.14}$$

To derive (2.13), we use (2.10) and (2.11):

$$\psi_k(xt) - \psi_k(t) = \sum_{i=0}^{k} (-1)^{k-i}(x^{p^{ni}} - x) \begin{bmatrix} k \\ i \end{bmatrix} t^{p^{ni}}$$

$$= \sum_{i=1}^{k} (-1)^{k-i}(x^{p^{nk}} - x) \begin{bmatrix} k-1 \\ i-1 \end{bmatrix}^{p^n} t^{p^{ni}}$$

$$= (x^{p^{nk}} - x)\psi_{k-1}^{p^n}(t).$$

To prove the second formula we make use of (2.12):

$$\sum_{i=0}^{k} (-1)^{k-i} \begin{bmatrix} k \\ i \end{bmatrix} t^{p^{ni}} = \sum_{i=0}^{k} (-1)^{k-i} t^{p^{ni}} \left\{ \begin{bmatrix} k-1 \\ i-1 \end{bmatrix}^{p^n} + F_{k-1}^{p^n-1} \begin{bmatrix} k-1 \\ i \end{bmatrix} \right\}$$

$$= \left\{ \sum_{i=1}^{k} (-1)^{k-i} \begin{bmatrix} k-1 \\ i-1 \end{bmatrix} t^{p^{n(i-1)}} \right\}^{p^n} + F_{k-1}^{p^n-1} \sum_{i=0}^{k-1} (-1)^{k-i} \begin{bmatrix} k-1 \\ i \end{bmatrix} t^{p^{ni}}$$

$$= \psi_{k-1}^{p^n}(t) - F_{k-1}^{p^n-1} \cdot \psi_{k-1}(t).$$

It is of some interest to notice that (2.14) may be derived without the use of (2.12). Indeed, from the product definition,

$$\psi_{k+1}(t) = \prod_{(c)} (t + c_0 x^k + \cdots + c_k)$$

$$= \prod_{c_0} \psi_k(t + c_0 x^k) = \prod_{c} \{\psi_k(t) + c\psi_k(x^k)\}$$

$$\psi_{k+1}(t) = \psi_k^{p^n}(t) - \psi_k^{p^n-1}(x^k)\psi_k(t). \tag{2.15}$$

But from the definition of $\psi_k(t)$ it is evident that $\psi_k(x^k) = F_k$, and hence (2.15) reduces to (2.14).

By means of the second method, (2.14) may be considerably extended. However, as the final formulas are not needed in the present paper, we shall not take the space to develop them here.

For later purposes it is useful to rewrite the product in (2.10). We group together the $p^n - 1$ polynomials cE, $c \neq 0$; then, since

$$\prod_{c \neq 0} (t - cE) = t^{p^n-1} - E^{p^n-1},$$

it follows that

$$\psi_k(t) = t \prod_{\deg G < k}' (t^{p^n-1} - G^{p^n-1}), \tag{2.16}$$

the product in this instance extending over the primary polynomials G of degree $< k$.

Again, if in (2.10) we put $t + x^k$ in place of t, we have as a second variant of (2.10) the formula useful later:

$$\psi_k(t) + F_k = \prod_{\deg G = k}' (t + G), \tag{2.17}$$

the product now extending over all primary G of degree $= k$.

3. The operator Δ. Formula (2.14) suggests the use of the operator Δ defined by

$$\Delta f(t) = f(xt) - xf(t), \tag{3.01}$$

where $f(t)$ is, to begin with, any linear polynomial $\sum_{i=1}^{k} A_i t^{p^{ni}}$. Thus (2.14) becomes

$$\Delta\psi_k(t) = (x^{p^{nk}} - x)\psi_{k-1}^{p^n}(t). \tag{3.02}$$

This formula in turn suggests that we define Δ^2 by means of

$$\Delta^2 f(t) = \Delta f(xt) - x^{p^n}\Delta f(t),$$

so that by (3.02)

$$\Delta^2\psi_k(t) = (x^{p^{nk}} - x)(x^{p^{n(k-1)}} - x)^{p^n}\psi_{k-2}^{p^{2n}}(t).$$

(Note that $\Delta^2 \neq \Delta\cdot\Delta$.) In general, we shall define

$$\Delta^{h+1} f(t) = \Delta^h f(xt) - x^{p^{nh}}\Delta^h f(t). \tag{3.03}$$

In particular, for the polynomial $\psi_k(t)$ it is clear that

$$\Delta^h\psi_k(t) = \frac{F_k}{F_{k-h}^{p^{nh}}}\psi_{k-h}^{p^{nh}}(t). \tag{3.04}$$

As an application of the last formula, consider any linear polynomial

$$g(t) = \sum_{j=0}^{k} A_j' t^{p^{nj}};$$

from the form of $\psi_k(t)$—in particular since the coefficient of the highest power of t is 1—it is clear that $g(t)$ may be put in the form

$$\sum_{j=0}^{k} A_j\psi_j(t). \tag{3.05}$$

We wish to determine the coefficients A_j. More generally, in place of (3.05) we write

$$g(tu) = \sum_{j=0}^{k} A_j(u)\psi_j(t). \tag{3.06}$$

If we put $t = 1$ in this formula, then, since $\psi_0(1) = 1$ and $\psi_j(1) = 0$ for $j > 0$, it is clear that $A_0(u) = g(u)$. Applying Δ to both members of (3.06), and using (3.02), we have

$$\Delta g(tu) = \sum_{j=1}^{k} A_j(u)(x^{p^{nj}} - x)\psi_{j-1}^{p^n}(t). \tag{3.07}$$

Again, put $t = 1$, and we see that $A_1(u) = g(u)/(x^{p^n} - x)$. To determine the general coefficient in (3.06), apply Δ^h to both members of the equality, and make use of (3.04); we find that

$$\Delta^h g(tu) = \sum_{j=h}^{k} A_j(u) \frac{F_j}{F_{j-h}^{p^{nh}}} \psi_{j-h}^{p^{nh}}(t).$$

But for $t = 1$, this implies

$$A_h(u) = \frac{\Delta^h g(u)}{F_h},$$

and therefore (3.06) becomes

$$g(tu) = \sum_{j=0}^{k} \frac{1}{F_j} \Delta^j g(u)\psi_j(t). \tag{3.08}$$

We state the

THEOREM 3.1. *Every linear polynomial $g(t)$ has a unique ψ-expansion* (3.08).

As an illustration of the expansion, let $g(u) = u^{p^{nk}}$, so that

$$\Delta g(u) = (x^{p^{nk}} - x)u^{p^{nk}},$$

$$\Delta^2 g(u) = (x^{p^{nk}} - x)(x^{p^{n(k-1)}} - x)^{p^n} u^{p^{nk}},$$

and generally for $j \leqq k$,

$$\Delta^j g(u) = \frac{F_k}{F_{k-j}^{p^{nj}}} u^{p^{nk}}.$$

Substitution in (3.08) leads to the identity

$$t^{p^{nk}} = \sum_{j=0}^{k} \frac{F_k}{F_j F_{k-j}^{p^{nj}}} \psi_j(t). \tag{3.09}$$

In exactly the same way, if we let $g(u) = \psi_k(u)$, and use (3.04), we are led to the identity

$$\psi_k(tu) = \sum_{j=0}^{k} \frac{F_k}{F_j F_{k-j}^{p^{nj}}} \psi_j(t)\psi_{k-j}^{p^{nj}}(u). \tag{3.10}$$

Returning to the beginning of this section, suppose we seek to generalize (3.02); more precisely, we wish to determine all $g_k(t) = \sum_{j=0}^{k} A_j^{(k)} t^{p^{nj}}$ such that

$$\Delta g_k(t) = G_k g_{k-1}^{p^n}(t), \tag{3.11}$$

where G_k does not involve t. Evidently (3.11) implies

$$\sum_{j=1}^{k} A_j^{(k)}(x^{p^{nj}} - x)t^{p^{nj}} = G_k \sum_{j=0}^{k-1} \{A_j^{(k-1)}\}^{p^n} t^{p^{n(j+1)}},$$

from which it follows that

$$A_j^{(k)} = \frac{G_k}{x^{p^{nj}} - x} \{A_{j-1}^{(k-1)}\}^{p^n}.$$

Repeated application of this formula leads to

$$A_j^{(k)} = \frac{\gamma_k}{\gamma_{k-j}^{p^{nj}}} \frac{\alpha_{k-j}^{p^{nj}}}{F_j},$$

where for brevity we write

$$\gamma_k = G_k G_{k-1}^{p^n} \cdots G_1^{p^{n(k-1)}},$$

and

$$\alpha_k = A_0^{(k)}.$$

Finally, therefore, any $g_k(t)$ satisfying (3.11) is of the form

$$\sum_{j=0}^{k} \frac{\gamma_k}{F_j} \left(\frac{\alpha_{k-j}}{\gamma_{k-j}} t \right)^{p^{nj}}. \tag{3.12}$$

Conversely it may be verified directly that (3.12) actually satisfies the condition (3.11). In particular, for $G_k = x^{p^{nk}} - x$, that is, $\gamma_k = F_k$, and

$$\alpha_k = (-1)^k F_k^k / L_k,$$

the polynomial (3.12) reduces to $\psi_k(t)$. If now we combine (3.11) with (3.08), we obtain the ψ-expansion

$$g_k(tu) = \sum_{j=0}^{k} \frac{\gamma_k}{\gamma_{k-j}^{p^{nj}}} \frac{g_{k-j}^{p^{nj}}(u)}{F_j} \psi_j(t), \tag{3.13}$$

a direct generalization of (3.10). We may state the

Theorem 3.2. *A set of linear polynomials $g_k(t)$ satisfying* (3.11) *is necessarily of the form* (3.12), *and conversely; $g_k(t)$ has the ψ-expansion* (3.13).

4. **The extended domain.** Following Artin,[5] we assign to x the "absolute value" $|x| = p^n$, and generally we put

$$|E| = p^{nk} \qquad \text{for deg } E = k,$$

$$|c| = 1 \qquad \text{for } c \text{ in } GF(p^n).$$

If further we define $|E/F| = |E|/|F|$, it is evident that

$$(4.01) \qquad \begin{aligned} |AB| &= |A|\cdot|B|, \\ |A+B| &\leqq \max(|A|, |B|), \end{aligned}$$

for all A, B rational in x. An expansion of the type

$$(4.02) \qquad \eta = c_k x^k + \cdots + c_0 + \frac{c_{-1}}{x} + \frac{c_{-2}}{x^2} + \cdots$$

is convergent for all sets of coefficients c_i in the Galois field. As a matter of fact, the series of absolute values is

$$p^{nk} + \cdots + 1 + p^{-n} + p^{-2n} = p^{nk}/(1 - p^{-n}).$$

More generally, we may define (convergent) expansions like (4.02) in which x is replaced by the symbol $x^{1/m}$, for m any positive integer,

$$(4.03) \qquad \alpha = c_k x^{k/m} + \cdots + c_0 + c_{-1}x^{-1/m} + \cdots .$$

The totality of expansions α evidently form a field $\mathfrak{F}(x, p^n)$. It will be convenient to call k the degree of η, and $|\eta| = p^{nk}$. More generally, we put $\deg \alpha = k/m$, $|\alpha| = p^{nk/m}$. If η_k be an η of degree k, then clearly every expansion

$$(4.04) \qquad \eta' = c_k\eta_k + \cdots + c_1\eta_1 + c_0 + c_{-1}\eta_{-1} + \cdots$$

is convergent; further, it is easily seen that it may be written in the form (4.02), that is, η' is also in $\mathfrak{F}$. The same remark applies to the expansions

$$\alpha' = c_k\alpha_k + \cdots + c_0 + c_{-1}\alpha_{-1} + \cdots .$$

It might appear that every algebraic equation

$$t^k + \alpha_1 t^{k-1} + \cdots + \alpha_k = 0, \qquad \alpha_i \text{ in } \mathfrak{F},$$

would be solvable in some $\mathfrak{F}(x, p^{nf})$. This is, however, not true. Thus, while the equation

$$t^{p^n} - t = \frac{1}{x}$$

has the solutions

$$t = c - x^{-1} + x^{-p^n} + x^{-p^{2n}} + \cdots, \qquad c \text{ in } GF(p^n);$$

[5] E. Artin, Mathematische Zeitschrift, vol. 19 (1924), pp. 153–246.

on the other hand

$$t^{p^n} - t = x \tag{4.05}$$

is solvable in no $\mathfrak{F}(x, p^{nf})$. For the sequel equations like (4.05) are of some interest; we introduce a symbol y,

$$\deg y = \frac{1}{p^n}, \qquad y^{p^n} - y = x;$$

all the solutions of (4.05) are furnished by $y + c$.

We next define the linear function in $\mathfrak{F}$

$$f(t) = \alpha_0 t + \alpha_1 t^{p^n} + \alpha_2 t^{p^{2n}} + \cdots; \tag{4.06}$$

note that

$$f(ct) = cf(t), \qquad f(t + u) = f(t) + f(u).$$

The series (4.06) may converge for all t in $\mathfrak{F}$, for example,

$$\sum_{k=0}^{\infty} c_k t^{p^{nk}}/F_k;$$

on the other hand

$$\sum_{k=0}^{\infty} c_k t^{p^{nk}}/L_k$$

converges only for $\deg t \leqq 1$.

Assuming the convergence of all the series involved, we define for $f(t)$ as in (4.06):

$$\Delta f(t) = f(xt) - xf(t) = \sum_{i=0}^{\infty} \alpha_i(x^{p^{ni}} - x)t^{p^{ni}}. \tag{4.07}$$

Consider now the ψ-expansion

$$f(tu) = \sum_{j=0}^{\infty} \beta_j(u)\psi_j(t). \tag{4.08}$$

Then exactly as in §3, we find

$$\beta_j(u) = \frac{\Delta^j f(u)}{F_j},$$

so that (4.08) becomes

$$f(tu) = \sum_{j=0}^{\infty} \frac{1}{F_j} \Delta^j f(u)\psi_j(t). \tag{4.09}$$

It may be shown without difficulty that (4.09) is valid for all t provided

$$|\Delta^j f(u)| \leqq |x^{p^{nj}}|.$$

Infinite products in $\mathfrak{F}$ involve no special difficulty. The product

$$\prod_{k=0}^{\infty} (1 + \alpha_k)$$

is convergent for $\Sigma\alpha_k$ convergent; for example, the product converges for deg $\alpha_k = -k$. In particular, we note for later purposes that the following products converge:

$$\lim_{k\to\infty} (1 + \alpha)^{(p^{nk}-1)/(p^n-1)} = \prod_{j=0}^{\infty} (1 + \alpha^{p^{nj}}), \qquad |\alpha| < 1;$$

$$(1 - x^{1-p^n})(1 - x^{1-p^{2n}})(1 - x^{1-p^{3n}}) \cdots .$$

5. **The fundamental function.** By (2.16) and (2.10)

$$t \prod_{\deg E<k}' (t^{p^n-1} - E^{p^n-1}) = \sum_{j=0}^{k} (-1)^{k-j} \begin{bmatrix} k \\ j \end{bmatrix} t^{p^{nj}}.$$

We divide both members of this equation by the product of the $(p^n - 1)^{\text{th}}$ powers of all primary E of degree $< k$, that is, by

$$(F_{k-1}F_{k-2} \cdots F_1)^{p^n-1} = \frac{F_k}{L_k};$$

the equality follows from (2.6). Thus

$$(5.01) \qquad t \prod_{\deg E<k}' \left(1 - \frac{t^{p^n-1}}{E^{p^n-1}}\right) = \sum_{j=0}^{k} (-1)^j \frac{L_k}{F_j L_{k-j}^{p^{nj}}} t^{p^{nj}}.$$

Suppose now we let k become infinite. The left member of (5.01) may be compared with

$$\prod_{k=0}^{\infty} \left(1 + \frac{|t|^{p^n-1}}{p^{nk(p^n-1)}}\right)^{p^{nk}},$$

which converges for all t in $\mathfrak{F}$ provided $p^n > 2$. In the case $p^n = 2$, it is necessary to group the factors in the infinite product in some manner that will insure (absolute) convergence; it is convenient to group the factors in the following way:

$$t \prod_{k=0}^{\infty} \prod_{\deg E=k}' \left(1 - \frac{t}{E}\right) \qquad (p^n = 2).$$

As for the right member of (5.01), we have formally at least

$$\sum_{j=0}^{\infty} (-1)^j \frac{t^{p^{nj}}}{F_j} \lim_{k\to\infty} \frac{L_k}{L_{k-j}^{p^{nj}}}.$$

Let us put

$$\xi_k = \frac{(x^{p^n} - x)^{(p^{nk}-1)/(p^n-1)}}{L_k}. \tag{5.02}$$

For k infinite, it follows from the end of §4 that ξ_k converges: we call the limit ξ_∞. Write

$$[k] = x^{p^{nk}} - x, \qquad [1]^{p^{nk}} = [k+1] - [k].$$

Then it is easily seen that (5.02) implies

$$\xi_k = \left(1 - \frac{[k-1]}{[k]}\right)\left(1 - \frac{[k-2]}{[k-1]}\right) \cdots \left(1 - \frac{[1]}{[2]}\right), \tag{5.03}$$

from which the convergence follows at once. Note that

$$|\xi_k| = |\xi_\infty| = 1.$$

Next, since by (5.03),

$$\xi_{k+1} - \xi_k = -\frac{[k]}{[k+1]}\xi_k,$$

we have

$$\deg(\xi_{k+1} - \xi_k) = \deg[k] - \deg[k+1] = -p^{nk}(p^n - 1);$$

and therefore, for $\delta_k = \xi_k - \xi_\infty$,

$$\deg \delta_k = -p^{nk}(p^n - 1). \tag{5.04}$$

It is now easy to prove that for k infinite the right side of (5.01) becomes

$$\sum_{j=0}^{\infty} \frac{(-1)^j}{F_j} t^{p^{nj}} \xi_\infty^{p^{nj}-1} X_j, \tag{5.05}$$

where for brevity we put

$$X_j = (x^{p^n} - x)^{(p^{nj}-1)/(p^n-1)}.$$

To derive (5.05), note that the right member of (5.01)

$$= \frac{1}{\xi_k} \sum_{j=0}^{k} \frac{(-1)^j}{F_j} t^{p^{nj}} X_j \xi_{k-j}^{p^{nj}} = \frac{1}{\xi_k} \sum_{j=0}^{k} \frac{(-1)^j}{F_j} t^{p^{nj}} X_j (\delta_{k-j}^{p^{nj}} + \xi_\infty^{p^{nj}})$$

To show that

$$\sum_{j=0}^{k} \frac{(-1)^j}{F_j} t^{p^{nj}} X_j \delta_{k-j}^{p^{nj}} \to 0,$$

break the sum into two parts:

$$\Sigma_1 = \sum_{2j<k}, \qquad \Sigma_2 = \sum_{2j \geq k};$$

$\Sigma_1 \to 0$ because of (5.04); $\Sigma_2 \to 0$ with

$$\sum_{2j \geqq k} \frac{(-1)^j}{F_j} t^{p^{nj}}.$$

It will be convenient to alter (5.05) slightly. If we put

$$z^{p^n-1} = x^{p^n} - x,$$

it is easily seen that z is in $\mathfrak{F}$; indeed, there is no difficulty in explicitly exhibiting z in the form (4.03). We now define

$$\xi = (x^{p^n} - x)^{1/(p^n-1)} \xi_\infty. \tag{5.06}$$

Then by (5.01) and (5.05) we have finally

$$t \prod_E{}' \left(1 - \frac{t^{p^n-1}}{E^{p^n-1}}\right) = \frac{1}{\xi} \sum_{j=0}^{\infty} \frac{(-1)^j}{F_j} t^{p^{nj}} \xi^{p^{nj}}, \tag{5.07}$$

the product extending over all primary E. If we put

$$\psi(t) = \sum_{j=0}^{\infty} \frac{(-1)^j}{F_j} t^{p^{nj}}, \tag{5.08}$$

we may state

THEOREM 5.1. *The linear function $\psi(t)$ defined by* (5.08) *for all t in $\mathfrak{F}$ vanishes for all ξE, E being an arbitrary polynomial, and for no other values of t. $\psi(t)$ has the period ξ:*

$$\psi(t + \xi E) = \psi(t).$$

That $\psi(t)$ has no zeros but ξE is evident from (5.07).

We next consider $\Delta\psi(t)$. From the definition of Δ it follows that

$$\begin{aligned}
\psi(xt) - x\psi(t) &= \sum_{j=0}^{\infty} \frac{(-1)^j}{F_j} (x^{p^{nj}} - x) t^{p^{nj}} \\
&= \sum_{j=1}^{\infty} \frac{(-1)^j}{F_{j-1}^{p^n}} t^{p^{nj}} = -\psi^{p^{nj}}(j);
\end{aligned}$$

that is,

$$\Delta\psi(t) = -\psi^{p^n}(t). \tag{5.09}$$

Generally, the definition of Δ^k in (3.03) leads to

$$\Delta^k\psi(t) = (-1)^k \psi^{p^{nk}}(t).$$

Therefore by (4.09) we have the formal ψ-expansion

$$\psi(tu) = \sum_{j=0}^{\infty} \frac{(-1)^j}{F_j} \psi^{p^{nj}}(u)\psi_j(t). \tag{5.10}$$

By the remark following (4.09), we may assert that (5.10) holds for all t, provided $|u| \leqq |x|$.

On the other hand, if t is a polynomial M, the series on the right side breaks off after a finite number of terms, and (5.10) holds for all u. Let the degree of M be m, then, since $\psi_j(M) = 0$ for $j > m$, (5.10) becomes

$$\psi(Mu) = \sum_{j=0}^{m} \frac{(-1)^j}{F_j} \psi_j(M) \psi^{p^{nj}}(u) . \tag{5.11}$$

This formula evidently exhibits $\psi(Mu)$ as a linear polynomial in $\psi(u)$:

$$\psi(Mu) = \omega_M(\psi(u)) . \tag{5.12}$$

We now factor the polynomial $\omega_M(t)$. For E an arbitrary polynomial, and $u = E\xi/M$, $\psi(Mu) = 0$; in other words,

$$\omega_M(t) = 0 \quad \text{for} \quad t = \psi\left(\frac{E}{M}\xi\right).$$

If then we restrict E to a complete residue system (mod M), say the set of all polynomials of degree $< m$, the quantities $\psi(E\xi/M)$ are all distinct. Further, they are exactly p^{nm} in number, whence the factorization

$$\omega_M(t) = A \prod_{E \pmod{M}} \left\{t - \psi\left(\frac{E}{M}\xi\right)\right\},$$

where A is independent of t. Making use of (5.11), it is readily seen that $A = (-1)^j \cdot$coefficient of x^m in M. We have therefore

THEOREM 5.2. *For $M = cx^m + \cdots + c_m$, $\omega_M(t)$ defined by* (5.12) *has the factorization*

$$\omega_M(t) = ct \prod_{\deg E < m}{}' \left\{\psi^{p^n-1}\left(\frac{E}{M}\xi\right) - t^{p^n-1}\right\}, \tag{5.13}$$

the product extending over primary E of degree $< m$.

The "multiplication problem" for the function $\psi(t)$ is solved by (5.11) and (5.13). The converse problem, that is, the algebraic problem of determining $\psi(t/M)$ in terms of $\psi(t)$, we leave for another paper. We remark that the group of the division problem is abelian for $t = \xi$ as well as for arbitrary t.

6. **The inverse linear function.** If $f(t)$ and $g(t)$ are linear functions

$$f(t) = \sum_{i=0}^{\infty} \alpha_i t^{p^{ni}}, \qquad g(t) = \sum_{i=0}^{\infty} \gamma_i t^{p^{ni}},$$

it is evident that $f(g(t))$ is also linear

$$f(g(t)) = \sum_{i=0}^{\infty} \gamma_i t^{p^{ni}},$$

where

$$\gamma_k = \sum_{k=i+j} \alpha_i \beta_j^{p^{ni}}. \tag{6.01}$$

In order to avoid the irrelevant question of convergence, we shall deal with the sets of coefficients

$$\{\alpha\} = \{\alpha_0, \alpha_1, \alpha_2, \cdots\}.$$

The product of two sets $\{\alpha\}$, $\{\beta\}$ is another set $\{\gamma\} = \{\gamma_0, \gamma_1, \gamma_2, \cdots\}$, where γ_k is determined by (6.01); we write

$$\{\gamma\} = \{\alpha\}\cdot\{\beta\}. \tag{6.02}$$

Note that in general the factors may not be permuted. The particular set $\{1, 0, 0, \cdots\}$ will be denoted by 1. It has the property

$$\{\alpha\}\cdot 1 = 1\cdot\{\alpha\} = \{\alpha\},$$

for all sets $\{\alpha\}$. We shall assume hereafter that α_0, the first element in any set $\{\alpha\}$, is different from zero. Then for given $\{\beta\}$, $\{\gamma\}$, the set $\{\alpha\}$ is unique, as is evident from (6.01); similarly $\{\beta\}$ is uniquely determined by $\{\alpha\}$ and $\{\gamma\}$.

In particular, we shall be interested in the case $\{\alpha\}\cdot\{\beta\} = 1$. We now prove the

THEOREM 6.1. *If $\{\alpha\}$, $\{\beta\}$ are two sets of coefficients such that $\{\alpha\}\cdot\{\beta\} = 1$, then also $\{\beta\}\cdot\{\alpha\} = 1$.*

The hypothesis of the theorem may by (6.01) be written thus

$$\sum_{k=i+j} \alpha_i \beta_j^{p^{ni}} = \begin{cases} 0 & \text{for } k > 0, \\ 1 & \text{for } k = 0. \end{cases} \tag{6.03}$$

We are to prove

$$\sum_{k=i+j} \beta_j \alpha_i^{p^{nj}} = \begin{cases} 0 & \text{for } k > 0, \\ 1 & \text{for } k = 0. \end{cases} \tag{6.04}$$

For $k = 0$ or 1, (6.03) and (6.04) are seen to be identical. We therefore assume (6.04) true for all $k < m$. Then by (6.03),

$$\begin{aligned} \sum_{h=0}^{m-1} \beta_h \alpha_{m-h}^{p^{nh}} &= -\alpha_0 \sum_{h=0}^{m-1} \beta_h \sum_{\substack{m-h=i+j \\ j>0}} \alpha_i^{p^{nh}} \beta_j^{p^{n(h+i)}} \\ &= -\alpha_0 \sum_{\substack{m=h+i+j \\ j>0}} \beta_h \alpha_i^{p^{nh}} \beta_j^{p^{n(h+i)}} \\ &= -\alpha_0 \sum_{\substack{m=k+j \\ j=0}} \left\{ \sum_{k=h+i} \beta_h \alpha_i^{p^{nh}} \right\} \beta_j^{p^{nk}}. \end{aligned}$$

Now since $j > 0$, $k < m$, and (6.04) may be applied to the inner sum, we have at once

$$\sum_{h=0}^{m-1} \beta_h \alpha_{m-h}^{p^{nh}} = -\alpha_0 \beta_m,$$

which shows that (6.04) holds for $k = m$. This complete the proof.

It is not difficult to solve (6.04) explicitly for $\{\beta\}$; however, the general form of the solution seems to be of no great interest. In the next section we shall find simple explicit solutions in certain particular cases by special methods.

7. **The inverse of $\psi(t)$ and $\psi_k(t)$.** We shall denote the inverse of $\psi(t)$ by $\lambda(t)$; we prove

THEOREM 7.1. *The inverse of $\psi(t)$ is*

$$\lambda(t) = \sum_{j=0}^{\infty} \frac{1}{L_j} t^{p^{nj}}, \tag{7.01}$$

defined for all $|t| \leqq |x|$.

In the notation of §6, the theorem becomes

$$\left\{1, -\frac{1}{F_1}, \frac{1}{F_2}, \cdots\right\} \cdot \left\{1, \frac{1}{L_1}, \frac{1}{L_2}, \cdots\right\} = 1. \tag{7.02}$$

By (6.03), (7.02) is equivalent to

$$\sum_{k=i+j} \frac{(-1)^i}{F_i L_j^{p^{ni}}} = \begin{cases} 0 & \text{for} \quad k > 0, \\ 1 & \text{for} \quad k = 0. \end{cases} \tag{7.03}$$

But (7.03) is an immediate consequence of Theorem 2.1. This proves Theorem 7.1. Application of Theorem 6.1 implies

$$\sum_{k=i+j} \frac{(-1)^i}{L_i F_j^{p^{ni}}} = \begin{cases} 0 \text{ for } k > 0, \\ 1 \text{ for } k = 0. \end{cases} \tag{7.04}$$

This also is implicit in an earlier result, namely, the identity (3.09).

As for the inverse of $\psi_k(t)$, it is somewhat more convenient to use the closely related polynomial

$$\pi_k(t) = (-1)^k \frac{L_k}{F_k} \psi_k(t). \tag{7.05}$$

We shall prove

THEOREM 7.2. *The inverse of $\pi_k(t)$ is*

$$\lambda_k(t) = \sum_{j=0}^{\infty} t^{p^{nj}} \frac{L_{k+j-1}}{L_j L_{k-1}^{p^{nj}}} \qquad (k \geqq 1), \tag{7.06}$$

defined for $|t| < 1$. *For* $k = 0$, $\lambda_0(t) = t$.

In the proof we shall for brevity deal with $\lambda_k(t)$ rather than its coefficient set. For $k = 0$, the theorem is obvious; for $k = 1$,

$$\pi_1(t) = t - t^{p^n}, \qquad \lambda_1(t) = t + t^{p^n} + t^{p^{2n}} + \cdots,$$

and it is evident that

$$\pi_1(\lambda_1(t)) = t,$$

so that our theorem holds for $k = 1$.

In order to carry out the induction, we shall need the following identities:

$$\pi_k(t) - \frac{1}{L_k^{p^n-1}} \pi_k^{p^n}(t) = \pi_{k+1}(t), \tag{7.07}$$

$$\lambda_k(tL_k) = \lambda_{k+1}(tL_k) - \lambda_{k+1}(t^{p^n}L_k). \tag{7.08}$$

The first formula follows directly from (2.14) and (7.05). The second is easily proved by substitution in (7.06).

We now assume our theorem true for k:

$$\pi_k(\lambda_k(t)) = t.$$

Then, on the one hand, the expression

$$\pi_k\{\lambda_k(tL_k)\} - \frac{1}{L_k^{p^n-1}} \pi_k^{p^n}\{\lambda_k(tL_k)\} = (t - t^{p^n})L_k.$$

But, on the other hand, it also

$$\begin{aligned} &= \pi_{k+1}\{\lambda_k(tL_k)\} && \text{by (7.07)},\\ &= \pi_{k+1}\{\lambda_{k+1}[(t - t^{p^n})L_k]\} && \text{by (7.08)}; \end{aligned}$$

and therefore

$$\pi_{k+1}\{\lambda_{k+1}(u)\} = u,$$

which completes the induction.

Returning to the function $\lambda(t)$ defined by (7.01), we note that it satisfies the relation

$$\lambda(xt) - x\lambda(t) = \lambda(t^{p^n}), \tag{7.09}$$

provided $|t| \leqq 1$. (7.09) may be at once verified by direct substitutions in (7.01). The formula is equivalent to (5.09).

We now ask whether it is possible to define $\lambda(t)$ for $\deg t > 1$. Obviously we shall require that the extended definition be such as to preserve the relation $\psi(\lambda(t)) = t$; this relation will at most determine $\lambda(t)$ to within an additive term $E\xi$; briefly $\lambda(t)$ will be determined (mod ξ). Now

$$\begin{aligned} \psi(\lambda(t + u)) = t + u &= \psi(\lambda(t)) + \psi(\lambda(u))\\ &= \psi\{\lambda(t) + \lambda(u)\} \end{aligned}$$

implies

$$\lambda(t + u) = \lambda(t) + \lambda(u) \pmod{\xi}; \tag{7.10}$$

similarly

$$\lambda(ct) = c\lambda(t) \pmod{\xi}. \tag{7.11}$$

We therefore assume (7.10), (7.11), and (7.09) in the form

$$\lambda(xt - t^{p^n}) = x\lambda(t) \pmod{\xi}. \tag{7.12}$$

Making use of (7.12) we may now define $\lambda(u)$ for $\deg u > 1$. For the equation

$$xt - t^{p^n} = u \tag{7.13}$$

has the solution

$$t = z(uz^{-p^n} + u^{p^n} z^{-p^{2n}} + u^{p^{2n}} z^{-p^{3n}} + \cdots), \qquad \deg t = \deg u - 1,$$

$$\text{for } z = x^{1/(p^n-1)}, \qquad \deg u < p^n/(p^n - 1).$$

On the other hand, for $\deg u \geqq p^n/(p^n - 1)$, the discussion of (4.04) indicates that, in a suitable extension of $\mathfrak{F}$, a symbol t may be defined satisfying (7.13) and such that $\deg t = \deg u/p^n$.

If then $\deg u < p^n$, we may define

$$\lambda(u) = x\lambda(t) \pmod{\xi}.$$

If $p^n \leqq \deg u < p^{2n}$, we use (7.13) and

$$xt_1 - t_1^{p^n} = t;$$

clearly $\deg t_1 < 1$, and we define

$$\lambda(u) = x\lambda(t) = x^2\lambda(t_1) \pmod{\xi}.$$

It is now evident how $\lambda(u)$ may be defined for all u.

THEOREM 7.3. *By using* (7.12) $\lambda(u)$ *may be defined for all* u; *the function satisfies* (7.10) *and* (7.11), *and furthermore* $\psi(\lambda(u)) = u$.

8. **The reciprocal of a linear function.** The reciprocal of a linear function is obviously not itself linear. In general, the coefficients of the reciprocal are rather complicated functions of the given coefficients; however, if the coefficients of the inverse function are assumed known, then certain coefficients of the reciprocal can be expressed very simply.

Let us use the notation

$$f(t) = \sum_{i=0}^{\infty} \alpha_i t^{p^{ni}} \qquad (\alpha_0 = 1);$$

$$g(t) = \sum_{i=0}^{\infty} \beta_i t^{p^{ni}},$$

the inverse of $f(t)$;

(8.01) $$h(t) = \frac{t}{f(t)} = \Sigma\, \gamma_m t^m ,$$

the reciprocal of $f(t)$. Evidently γ_m is zero unless m is a multiple of $p^n - 1$. From the definition of γ_m, we have at once

(8.02) $$\sum_{p^{ni} \leq m} \gamma_{m-p^{ni}}\, \alpha_i = \begin{cases} 0 \text{ for } m > 1 , \\ 1 \text{ for } m = 1 , \end{cases}$$

from which γ_m may be calculated recursively. We shall now show that for $m = p^{nk} - p^{ni}$, γ_m is of particularly simple form; indeed,

(8.03) $$\gamma_{p^{nk}-p^{ni}} = \beta_{k-i}^{p^{ni}} .$$

To begin with, we prove

(8.04) $$\gamma_{pm} = \gamma_m^p .$$

For consider

$$h(t) - h^p(t) = \frac{t}{f(t)} - \left(\frac{t}{f(t)}\right)^p = \left(\frac{t}{f(t)}\right)^p \left\{\left(\frac{f(t)}{t}\right)^{p-1} - 1\right\} .$$

Clearly

$$\left(\sum_{i=0}^{\infty} \alpha_i t^{p^{ni}-1}\right)^{p-1} - 1$$

has no term in t^{pm}, and therefore $h(t) - h^p(t)$ has no term in t^{pm}. On the other hand, by (8.01),

$$\begin{aligned} h(t) - h^p(t) &= \sum_{m=0}^{\infty} \gamma_m t^m - \sum_{m=0}^{\infty} \gamma_m^p t^{pm} \\ &= \sum_{p \nmid m} \gamma_m t^m + \sum_{\text{all } m} (\gamma_{pm} - \gamma_m^p)\, t^{pm} . \end{aligned}$$

The coefficient of every t^{pm} must vanish; that is,

$$\gamma_{pm} - \gamma_m^p = 0$$

which proves (8.04).

The proof of (8.03) is now quite simple. In view of (8.04), it is only necessary to show that

(8.05) $$\gamma_{p^{nk}-1} = \beta_k .$$

For $k = 0$, (8.05) reduces to $\gamma_0 = \beta_0$, which agrees with (8.02). Let us therefore assume that (8.05) holds up to and including $k - 1$. By (8.02), for $k > 0$,

$$\sum_{i \leq k} \gamma_{p^{nk}-p^{ni}}\, \alpha_i = 0 ,$$

or what amounts to the same thing

$$\gamma_{p^{nk}-1} = -\sum_{i=1}^{k} \gamma_{p^{nk}-p^{ni}}\,\alpha_i$$

$$= -\sum_{i=1}^{k} \gamma_{p^{n(k-i)}-1}^{p^{ni}}\,\alpha_i \qquad \text{(by (8.04))}$$

$$(8.06) \qquad = -\sum_{i=1}^{k} \beta_{k-i}^{p^{ni}}\,\alpha_i\,,$$

by our assumption on (8.05). On the other hand, making use of (6.03), we have at once

$$\beta_k = -\sum_{i=1}^{k} \beta_{k-i}^{p^{ni}}\,\alpha_i\,.$$

Comparison with (8.06) shows that

$$\gamma_{p^{nk}-1} = \beta_k\,,$$

which completes the induction. We have therefore

THEOREM 8.1. *If γ_m is the general coefficient of the reciprocal* (8.01) *and β_m is the general coefficient of the inverse; if further $m = p^{nk} - p^{ni}$, then*

$$\gamma_m = \beta_{k-i}^{p^{ni}}\,.$$

Let us consider next

$$(8.07) \qquad f^{p^{nj}-1}(t) = \left(\sum_{i=0}^{\infty} \alpha_i t^{p^{ni}}\right)^{p^{nj}-1}.$$

If the right member be expanded, the first term is evidently $t^{p^{nj}-1}$; we shall now show that the expansion contains no further term in $t^{p^{nk}-1}$ $(k > j)$. To prove this we write

$$f^{p^{nj}-1}(t) = \frac{f^{p^{nj}}(t)}{t}\,\frac{t}{f(t)}$$

$$(8.08) \qquad = \sum_{i=0}^{\infty} \alpha_i^{p^{nj}}\, t^{p^{n(i+j)}-1} \sum_{m=0}^{\infty} \gamma_m t^m\,.$$

The coefficient of $t^{p^{n(k+j)}-1}$ in the product (8.08)

$$= \sum_{p^{n(k+j)}=p^{n(i+j)}+m} \alpha_i^{p^{nj}}\,\gamma_m = \sum_{i=0}^{k} \alpha_i^{p^{nj}}\,\gamma_{p^{n(k+j)}-p^{n(i+j)}}$$

$$= \sum_{i=0}^{k} \alpha_i^{p^{nj}} \beta_{k-i}^{p^{n(i+j)}} \qquad \text{by (8.03)},$$

$$= 0 \qquad \text{for } k > 0\,, \quad \text{by (6.03)}\,.$$

We have therefore

THEOREM 8.2. *The* $(p^{nj}-1)^{\text{th}}$ *power of a linear function begins with a term in* $t^{p^{nj}-1}$; *it has no term in* $t^{p^{nk}-1}$ *for* $k > j$.

To obtain a better result concerning (8.07), note that if we put

$$f^{p^{nk}-1}(t) = \sum_{m=0}^{\infty} \gamma_m^{(k)} t^m, \tag{8.09}$$

then by (8.08),

$$\gamma_{m-1}^{(k)} = \sum_{m=p^{n(i+k)}+j} \alpha_i^{p^{nk}} \gamma_j .$$

In particular,

$$\begin{aligned}\gamma_{p^n m-1}^{(k)} &= \sum_{p^n m = p^{n(i+k)}+j} \alpha_i^{p^{nk}} \gamma_j \\ &= \sum_{m=p^{n(i+k-1)}+j} \alpha_i^{p^{nk}} \gamma_j^{p^n} \qquad \text{by (8.04)},\end{aligned}$$

and therefore

$$\gamma_{p^n m-1}^{(k)} = \left(\gamma_{m-1}^{(k-1)}\right)^{p^n}. \tag{8.10}$$

In particular,

$$\gamma_{p^n m-1}^{(1)} = \left(\gamma_{m-1}^{(0)}\right)^{p^n} = \begin{cases} 1 & \text{for} \quad m = 1, \\ 0 & \text{for} \quad m > 1, \end{cases}$$

and therefore by repeated application of (8.10) we have

$$\gamma_{p^{nk} m-1}^{(k)} = 0 \quad \text{for} \quad m > 1 .$$

Hence the generalization of Theorem 8.2:

THEOREM 8.3. *The* $(p^{nj}-1)^{\text{th}}$ *power of a linear function contains no term of exponent* $(p^{nj}m-1)$, *for* $m > 1$.

Finally, let us note the following connection between γ_m and $\gamma_m^{(k)}$.

$$h(t) = \frac{t}{f(t)} = \frac{g(f(t))}{f(t)} = \sum_{k=0}^{\infty} \beta_k f^{p^{nk}-1}(t),$$

and therefore by (8.01) and (8.09)

$$\gamma_m = \sum_{p^{nk} \leqq m+1} \beta_k \gamma_m^{(k)} . \tag{8.11}$$

9. **The reciprocal of $\psi(t)$ and $\psi_k(t)$.** For the particular linear functions $\psi(t)$ and $\psi_k(t)$ we define Γ_m and $\Gamma_{k,m}$ by means of

$$\frac{t}{\psi(t)} = \sum_{m=0}^{\infty} \Gamma_m t^m, \qquad \frac{t}{\pi_k(t)} = \sum_{m=0}^{\infty} \Gamma_{k,m} t^m, \tag{9.01}$$

from which follow the recursion formulas

$$\sum_{p^{ni} \leq m} \frac{(-1)^i}{F_i} \Gamma_{m-p^{ni}} = \begin{cases} 0 & \text{for} \quad m > 1, \\ 1 & \text{for} \quad m = 1; \end{cases}$$

$$\sum_{p^{ni} \leq m} (-1)^i \frac{L_k}{F_i L_{k-1}^{p^{ni}}} \Gamma_{k,m-p^{ni}} = \begin{cases} 0 & \text{for} \quad m > 1, \\ 1 & \text{for} \quad m = 1. \end{cases}$$

In general, Γ_m and $\Gamma_{k,m}$ are rather complicated. However, for $m = p^{n(i+j)} - p^{nj}$, the first theorem of §8 together with Theorems 7.1 and 7.2 imply the following

THEOREM 9.1. *For Γ_m, $\Gamma_{k,m}$ as defined by* (9.01); *and $m = p^{n(i+j)} - p^{nj}$, we have*

$$\Gamma_m = \frac{1}{L_i^{p^{nj}}}, \qquad \Gamma_{k,m} = \left(\frac{L_{k+i-1}}{L_i L_{k-1}^{p^{ni}}}\right)^{p^{nj}}. \tag{9.02}$$

As a generalization of the first of (9.01), put

$$\frac{\psi(tu)}{\psi(t)} = \sum_{m=0}^{\infty} \Gamma_m(u) t^m, \qquad \psi^{p^{nk}-1}(t) = \sum_{m=0}^{\infty} \Gamma_m^{(k)} t^m, \tag{9.03}$$

so that $\Gamma_m(u)$ is a polynomial in u. Indeed, since

$$\frac{\psi(tu)}{\psi(t)} = \frac{\psi(tu)}{t} \frac{t}{\psi(t)} = \sum_{i=0}^{\infty} \frac{(-1)^i}{F_i} u^{p^{ni}} t^{p^{ni}-1} \sum_{m=0}^{\infty} \gamma_m t^m,$$

it follows from the first of (9.03) that

$$\Gamma_{m-1}(u) = \sum_{p^{ni} \leq m} \frac{(-1)^i}{F_i} \gamma_{m-p^{ni}} u^{p^{ni}}. \tag{9.04}$$

On the other hand, by (5.10)

$$\frac{\psi(tu)}{\psi(t)} = \sum_{i=0}^{\infty} \frac{(-1)^i}{F_i} \psi_i(u) \psi^{p^{ni}-1}(t);$$

therefore, by the second of (9.03),

$$\Gamma_{m-1}(u) = \sum_{p^{ni} \leq m} \frac{(-1)^i}{F_i} \Gamma_{m-1}^{(i)} \psi_i(u). \tag{9.05}$$

In particular,

$$\Gamma_{p^{nk}-1}(u) = \frac{(-1)^k}{F_k} \psi_k(u).$$

Finally by (9.04) and (8.04), or by (9.05) and (8.10), we have

$$\Delta \Gamma_{p^n m-1}(u) = -\Gamma_{m-1}^{p^n}(u).$$

We now investigate the connection between the Γ's and certain power sums of polynomials. To begin with, we need a lemma.

If

$$f(t) = \prod_{i=1}^{k} (t + \alpha_i) = t^k + \beta_1 t^{k-1} + \cdots + \beta_k ,$$

$$f'(t) = kt^{k-1} + (k-1)\beta_1 t^{k-2} + \cdots + \beta_{k-1} ,$$

it may be proved, by induction on k, that

$$\frac{f'(t)}{f(t)} = \sum_{i=1}^{k} \frac{1}{t + \alpha_i} . \tag{9.06}$$

In this identity let us take $f(t) = \psi_k(t)$. Evidently $f'(t)$ reduces to $(-1)^k F_k/L_k$, and therefore by (2.10) and (7.05),

$$\frac{1}{\pi_k(t)} = \frac{(-1)^k}{\psi_k(t)} \frac{F_k}{L_k} = \sum_{\deg E < k} \frac{1}{t - E} , \tag{9.07}$$

the summation extending over all E (including 0) of degree $< k$. Similarly, if we take $f(t) = \psi_k(t) + F_k$, $f'(t)$ again reduces to $(-1)^k F_k/L_k$, and therefore by (2.17),

$$\frac{(-1)^k}{\psi_k(t) + F_k} \frac{F_k}{L_k} = \sum_{\deg G = k}' \frac{1}{t + G} , \tag{9.08}$$

the summation now extending over all primary G of degree k. Put $t = 0$ and get the result

$$\sum_{\deg G = k}' \frac{1}{G} = \frac{(-1)^k}{L_k} . \tag{9.09}$$

It is now convenient for the applications to change slightly the right member of (9.07). By the identity (9.06), for $f(t) = t^{p^n-1} - u^{p^n-1}$,

$$\frac{-t^{p^n-2}}{t^{p^n-1} - u^{p^n-1}} = \sum_{c \neq 0} \frac{1}{t - cu}$$

summed over all $c \neq 0$ in $GF(p^n)$. If then we group together all E of equal degree, (9.07) becomes

$$\frac{1}{\pi_k(t)} = \frac{1}{t} + \sum_{\deg G < k}' \frac{-t^{p^n-2}}{t^{p^n-1} - G^{p^n-1}} , \tag{9.10}$$

the summation now extending over primary G only. Then for $|t|$ sufficiently small, (9.10) implies

$$\frac{t}{\pi_k(t)} = 1 + \sum_{m=1}^{\infty} t^{(p^n-1)m} \sum_{\deg G = m}' G^{-(p^n-1)m} .$$

Comparison with the second of (9.01) leads at once to

THEOREM 9.2. *The sum of the reciprocals of the* m^{th} *powers of the primary polynomials of degree* $< k$ *is* $\Gamma_{k,m}$ *defined in* (9.01), *provided* m *is divisible by* $p^n - 1$. *In particular, by* (9.02),

$$\sum_{\deg G<k}{}' G^{-m} = \frac{L_{k+j-1}}{L_j L_{k-1}^{p^{nj}}} \qquad \text{for } m = p^{nj} - 1 . \tag{9.11}$$

COROLLARY. *The sum of the reciprocals of the* m^{th} *powers of the primary polynomials of degree* k *is* $\Gamma_{k+1,m} - \Gamma_{k,m}$, *provided* m *is a multiple of* $p^n - 1$.

It may be shown (exactly as in proving (5.05)) that for k infinite (9.10) becomes

$$\frac{t}{\psi(t)} = 1 + \sum_G{}' \frac{t^{p^n-1}}{(\xi E)^{p^n-1} - t^{p^n-1}}, \tag{9.12}$$

summed over all primary G. (For $p^n = 2$ see the remark at the beginning of §5.) For sufficiently small $|t|$, the right member of (9.12) may be expanded in the form

$$1 + \sum_{m=1}^{\infty} \left(\frac{t}{\xi}\right)^{(p^n-1)m} \sum_G{}' G^{-(p^n-1)m}.$$

Comparison with the first of (9.01) leads to

THEOREM 9.3. *The sum of the reciprocals of the* m^{th} *powers of all primary polynomials is* Γ_m, *defined in* (9.01), *provided* m *is divisible by* $p^n - 1$. *Finally by* (9.02),

$$\sum_G{}' G^{-m} = \frac{\xi^{p^{nj}-1}}{L_k} \qquad \text{for } m = p^{nj} - 1 . \tag{9.13}$$

It should be remarked that we may pass directly from Theorem 9.2 to a theorem equivalent to Theorem 9.3 without the use of (9.12). However, the method used here seems somewhat simpler.

By means of the identity (9.08) it is possible to evaluate $\Sigma' G^{-m}$, summed over primary G of degree k, in the case m not a multiple of $p^n - 1$. The simplest instance, $m = 1$, is given in (9.09). For general m it is necessary to expand the left member of (9.08) in a series of ascending powers of t. We put

$$\frac{F_k}{F_k + \psi_k(t)} = \left\{1 + (-1)^k \sum_{i=0}^{k} \frac{(-1)^i t^{p^{ni}}}{F_i L_{k-i}^{p^{ni}}}\right\}^{-1} = \Sigma \Lambda_m^{(k)} t^m. \tag{9.14}$$

Note in particular that

$$\Lambda_m^{(k)} = (-1)^{(k-1)m} L_k^{-m} \quad \text{for} \quad m \leqq p^n - 1.$$

If next we expand the right member of (9.08):

$$\sum_{\deg G=k}' \frac{1}{t+G} = \sum_{m=0}^{\infty} t^m \sum_{\deg G=k}' \frac{1}{G^{m+1}};$$

comparison with (9.14) yields

THEOREM 9.4. *For m an arbitrary positive integer, the sum of the reciprocals of the m^{th} powers of the primary polynomials of degree k is $(-1)^k \Lambda_{m-1}^{(k)}/L_k$, where $\Lambda_{m-1}^{(k)}$ is defined by* (9.10). *In particular*

$$\sum_{\deg G=k}' \frac{1}{G^m} = -\frac{(-1)^{(k-1)m}}{L_k^m} \quad \text{for} \quad 1 \leqq m \leqq p^n.$$

It is also possible, by use of (9.08), to evaluate $\Sigma' G^m$, summed over primary G of degree k, for $m > 0$. We expand in series of *descending* powers of t. Put

$$\frac{1}{\psi_k(t) + F_k} = t^{-p^{nk}}\left\{1 - \begin{bmatrix} k \\ k-1 \end{bmatrix} t^{-p^{n(k-1)}(p^n-1)} + \begin{bmatrix} k \\ k-2 \end{bmatrix} t^{-p^{n(k-2)}(p^{2n}-1)} - \cdots\right\}^{-1}$$

$$= \sum_{m=p^{nk}}^{\infty} H_m^{(k)} t^{-m}. \tag{9.15}$$

But we have for the right member of (9.08)

$$\sum_{\deg G=k}' \frac{1}{t+G} = \sum_{m=1}^{\infty} (-1)^{m-1} t^{-m} \sum_{\deg G=k}' G^{m-1}.$$

Comparison with (9.15) leads to the

THEOREM 9.5. *For such m as appear in* (9.15), *the sum of the $(m-1)^{\text{th}}$ powers of the primary polynomials of degree k is $(-1)^{k+m-1} H_m^{(k)} F_k/L_k$. In particular*

$$\sum_{\deg G=k}' G^{p^{nk}-1} = (-1)^k \frac{F_k}{L_k}.$$

10. **Binomial congruences.** Consider the congruence

$$t^{p^n-1} \equiv A \pmod{P}, \qquad P \nmid A, \tag{10.01}$$

where P is irreducible of degree k. Define the symbol $\{A/P\}$ as that element in $GF(p^n)$ such that

$$\left\{\frac{A}{P}\right\} \equiv A^{(p^{nk}-1)/(p^n-1)} \pmod{P}.$$

A necessary and sufficient condition that the congruence (10.01) be solvable is $\{A/P\} = 1$. We have the following theorem of higher reciprocity due to F. K. Schmidt:[6]

THEOREM 10.1. *If P and Q are distinct primary irreducible polynomials of degree k and l respectively, then*

$$\left\{\frac{P}{Q}\right\} = (-1)^{kl}\left\{\frac{Q}{P}\right\}. \tag{10.02}$$

We shall give a new proof of this theorem analogous to Eisenstein's proofs of the ordinary theorems of quadratic and biquadratic reciprocity. The proof depends on the analog of Gauss' Lemma:[7]

$$\left\{\frac{A}{P}\right\} = \prod_{\deg E<k}{}' \operatorname{sgn} \mathscr{R}\left(\frac{AE}{P}\right), \tag{10.03}$$

where $\mathscr{R}(A/M)$ denotes the remainder in the division of A by M, sgn E is the coefficient of the highest power of x occurring in E, and the product is extended over the $(p^{nk} - 1)/(p^n - 1)$ primary polynomials of degree $< k$.

We shall now express $\{A/P\}$ in terms of the ψ-function. Note that

$$AE \equiv bB \pmod P \quad \text{implies} \quad \psi\left(\frac{AE}{P}\xi\right) = b\psi\left(\frac{B}{P}\xi\right),$$

and from this follows, for $P \nmid A$,

$$\prod_{\deg E<k}{}' \frac{\psi\left(\dfrac{AE}{P}\xi\right)}{\psi\left(\dfrac{E}{P}\xi\right)} = \prod_{\deg E<k}{}' \operatorname{sgn} \mathscr{R}\left(\frac{AE}{P}\right) = \left\{\frac{A}{P}\right\}, \tag{10.04}$$

by Gauss' Lemma (10.03).

Now by (5.10) and (5.11), for primary A,

$$\frac{\psi\left(\dfrac{AE}{P}\xi\right)}{\psi\left(\dfrac{E}{P}\xi\right)} = \prod_{\deg F<\deg A}{}' \left\{\psi^{p^n-1}\left(\frac{F}{A}\xi\right) - \psi^{p^n-1}\left(\frac{E}{P}\xi\right)\right\},$$

so that we have by (10.04),

$$\left\{\frac{A}{P}\right\} = \prod_{E,F}{}' \left\{\psi^{p^n-1}\left(\frac{F}{A}\xi\right) - \psi^{p^n-1}\left(\frac{E}{P}\xi\right)\right\},$$

[6] F. K. Schmidt, Erlanger Sitzungsberichte, vols. 58–59 (1928), pp. 159–172. I am indebted to Ore for pointing out Schmidt's priority. For other proofs of the theorem see my papers, American Journal of Mathematics, vol. 54 (1932), pp. 39–50; Bulletin of the American Mathematical Society, vol. 39 (1933), pp. 155–160; O. Ore, Transactions of the American Mathematical Society, vol. 36 (1934), pp. 243–274.

[7] American Journal of Mathematics, vol. 54 (1932), pp. 39–50.

the product extending over primary E, F, $\deg E < \deg P$, $\deg F < \deg A$. If now we put $A = Q$, and note that for p odd,

$$\frac{p^{nk}-1}{p^n-1}\frac{p^{nl}-1}{p^n-1} \equiv kl \pmod{2},$$

(10.02) follows immediately.

It is not difficult by exactly the same method to prove the reciprocity theorem for any pair of polynomials having no factor in common; the only essential point is the extension of Gauss' Lemma.

11. **Trinomial congruences.** In this section we shall discuss some properties of the congruence

$$t^{p^n} - Bt \equiv A \pmod{P}, \tag{11.01}$$

where P is irreducible of degree k, and B is not divisible by P. We distinguish two cases: (i) $\{B/P\} \neq 1$; (ii) $\{B/P\} = 1$. Here $\{B/P\}$ is the symbol defined in §10.

Case (i) is easily disposed of. Assume (11.01) solvable. Raise both members of the congruence to the p^{nj}-th power:

$$t^{p^{n(j+1)}} - B^{p^{nj}}t^{p^{nj}} \equiv A^{p^{nj}} \pmod{P} \qquad (j = 0, \cdots, k-1).$$

Multiply both sides of this congruence by

$$B^{p^{n(k-1)}+p^{n(k-2)}+\cdots+p^{n(j+1)}},$$

and sum for $j = 0, 1, \cdots, k-1$: we find that

$$t^{p^{nk}} - B^{(p^{nk}-1)/(p^n-1)}t \equiv \sum_{j=0}^{k-1} A^{p^{nj}}B^{(p^{nk}-p^{n(j+1)})/(p^n-1)}. \tag{11.02}$$

But

$$t^{p^{nk}} \equiv t, \qquad B^{(p^{nk}-1)/(p^n-1)} \equiv \left\{\frac{B}{P}\right\},$$

so that (11.02) becomes

$$t \equiv \frac{\sum_{j=0}^{k-1} A^{p^{nj}}B^{(p^{nk}-p^{n(j+1)})/(p^n-1)}}{1-\{B/P\}}. \tag{11.03}$$

Conversely, by direct substitution it may be verified that this value of t satisfies (11.01). Therefore, we have the

THEOREM 11.1. *If* $\{B/P\} \neq 1$, *the congruence* (11.01) *always has the unique solution furnished by* (11.03).

Case (ii), $\{B/P\} = 1$. By §10, a polynomial C exists such that

$$B \equiv C^{p^n-1} \pmod{P}.$$

Clear $P \nmid C$. Then (11.01) may be written in the form

$$u^{p^n} - u \equiv M \qquad \pmod{P}, \tag{11.04}$$

where we have put

$$u \equiv Ct, \qquad M \equiv \frac{A}{C^{p^n}}.$$

We shall therefore confine ourselves to the congruence (11.04). Raising both members of the congruence to the p^{nj}-th power, and adding the resulting congruences, we arrive at the *necessary* condition.

$$M + M^{p^n} + \cdots + M^{p^{n(k-1)}} \equiv 0 \qquad \pmod{P}. \tag{11.05}$$

To show that (11.05) is also *sufficient* for the solvability of (11.04), note first that the left member of (11.05) is, for all M, congruent (mod P) to an element of $GF(p^n)$. This follows from the identity

$$M^{p^{nk}} - M \equiv \prod_c \{c + M + M^{p^n} + \cdots + M^{p^{n(k-1)}}\}, \tag{11.06}$$

the product extending over all c in $GF(p^n)$. We shall denote the residue by $\rho(M) = \rho(M, P)$:

$$\rho(M, P) \equiv M + M^{p^n} + \cdots + M^{p^{n(k-1)}} \qquad \pmod{P}. \tag{11.07}$$

For fixed P, the symbol $\rho(A)$ has the properties

$$\rho(A + B) = \rho(A) + \rho(B),$$
$$\rho(cA) = c\rho(A), \qquad c \text{ in } GF(p^n),$$
$$\rho(c) = kc;$$

all of these follow directly from (11.07). It is easily seen, by using (11.06), that for fixed c, $\rho(A) = c$ for exactly $p^{n(k-1)}$ incongruent A's (mod P); in particular, there are $p^{n(k-1)}$ incongruent A's such that $\rho(A) = 0$. On the other hand, if E run through a complete residue (mod P), $E^{p^n} - E$ takes on exactly $p^{n(k-1)}$ incongruent values; indeed

$$E^{p^n} - E \equiv F^{p^n} - F$$

implies

$$(E - F)^{p^n} \equiv E - F,$$

whence $E - F = c$, an element of the Galois field. This evidently establishes the sufficiency of (11.05).

THEOREM 11.2. *A necessary and sufficient condition that* (11.04) *be solvable is* $\rho(M, P) = 0$; *the symbol* $\rho(M, P)$ *is defined by* (11.07).

The criterion $\rho(M, P) = 0$ is not very satisfactory; it is roughly on a par with the Euler criterion for the congruence (10.01). We shall now establish the much better criterion mentioned in the Introduction:

THEOREM 11.3. *The congruence* (11.04) *is solvable if and only if*

$$MP' \equiv c_1 x^{k-2} + c_2 x^{k-3} + \cdots + c_{k-1} \pmod{P},$$

that is, to a polynomial of degree $< k-1$; *here* P' *denotes the derivative of* P *with respect to* x:

$$P = x^m + b_1 x^{m-1} + \cdots + b_m, \quad P' = mx^{m-1} + (m-1)b_1 x^{m-2} + \cdots + b_{m-1}.$$

The proof depends on the identity (5.01), which it will be recalled is a variant of (2.16). We shall find it convenient to give that identity the form

$$(11.08) \qquad t \prod_{\deg E < k-1} \left(1 - \frac{tL_{k-1}}{E}\right) = \sum_{i=0}^{k-1} (-1)^i \frac{L_{k-1}^{p^{ni}}}{F_i L_{k-1-i}^{p^{ni}}} t^{p^{ni}},$$

the product extending over all E (except 0) of degree $< k-1$. Now by (2.06),

$$\begin{aligned}\left(\frac{L_{k-1}}{L_{k-1-i}}\right)^{p^{ni}} &= \{(x^{p^{n(k-1)}} - x) \cdots (x^{p^{n(k-i)}} - x)\}^{p^{ni}} \\ &\equiv (-1)^i (x^{p^{ni}} - x)(x^{p^{ni}} - x^{p^n}) \cdots (x^{p^{ni}} - x^{p^{n(i-1)}}) \\ &\equiv (-1)^i F_i \pmod{P}.\end{aligned}$$

Substitution in (11.08) leads at once to the identical congruence

$$(11.09) \qquad t \prod_{\deg E < k-1} \left(1 - \frac{tL_{k-1}}{E}\right) \equiv \sum_{i=0}^{k-1} t^{p^{ni}} \pmod{P}.$$

We see therefore that $\rho(M, P) = 0$ for ML_{k-1} congruent (mod P) to a polynomial of degree $< k-1$. We now show that

$$(11.10) \qquad (-1)^{k-1} L_{k-1} \equiv P' \pmod{P}.$$

Let us use the fuller notation $P(x)$ in place of P. Then

$$(11.11) \qquad P(t) \equiv (t - x)(t - x^{p^n}) \cdots (t - x^{p^{n(k-1)}}) \pmod{P(x)}.$$

But as already noted in §9, if

$$t^k + a_1 t^{k-1} + \cdots + a_k = (t - b_1) \cdots (t - b_k),$$

then

$$kt^{k-1} + (k-1)a_1 t^{k-2} + \cdots + a_{k-1} = (t - b_1) \cdots (t - b_k) \sum_{i=1}^{k} \frac{1}{t - b_i}.$$

Therefore (11.11) implies

$$P'(t) \equiv P(t) \sum_{i=0}^{k-1} \frac{1}{t - x^{p^{ni}}},$$

and therefore

$$\begin{aligned} P'(x) &\equiv (x - x^{p^n})(x - x^{p^{2n}}) \cdots (x - x^{p^{n(k-1)}}) \\ &\equiv (-1)^{k-1} L_{k-1}, \end{aligned}$$

which is identical with (11.10). This completes the proof of the theorem.

THEOREM 11.4. *If the product*

$$MP' \equiv cx^{k-1} + c_1 x^{k-2} + \cdots + c_{k-1} \pmod{P}, \tag{11.12}$$

then $\rho(M, P) = c$.

For $c = 0$, our theorem reduces to the last. Assume, therefore, that $c \neq 0$. Making use of (11.10), rewrite (11.12) in the form

$$(-1)^{k-1} L_{k-1} M \equiv cG \pmod{P},$$

where G is primary of degree $k - 1$. The identity (11.09) implies

$$(-1)^{k-1} L_{k-1} \rho(M) \equiv cG \prod \left(1 - \frac{cG}{E}\right) \equiv cG \prod \frac{cG - E}{E}; \tag{11.13}$$

in both cases the product is taken over all E (except 0) of degree $< k - 1$. Since there are $p^{n(k-1)} - 1$ such E, we have

$$cG \prod_E (cG - E) = c^{p^{n(k-1)}} F_{k-1} = cF_{k-1},$$

$$\prod_E E = (-1)^{k-1} (F_1 F_2 \cdots F_{k-2})^{p^n - 1},$$

so that

$$cG \prod_E \frac{cG - E}{E} = (-1)^{k-1} cL_{k-1};$$

and therefore by comparison with (11.13)

$$\rho(M) = c.$$

This completes the proof of the theorem.

In order to generalize somewhat the congruence (11.4), note that

$$u^{p^n} - u \equiv A \quad \text{implies} \quad U^{p^e} - U \equiv A,$$

where

$$n = ef, \qquad U \equiv u + u^{p^e} + \cdots + u^{p^{e(f-1)}};$$

however, the converse is not true in general. We therefore consider the congruence

$$U^{p^e} - U \equiv A \pmod{P}, \qquad n = ef. \tag{11.14}$$

It is evident that if (11.14) is solvable, it has p^e solutions; indeed, if U_0 is one solution, then all are included in the formula $U_0 + a$ where a is in $GF(p^e)$.

As a generalization of Theorem 11.2, we may prove *that*

$$(11.15) \qquad A + A^{p^e} + \cdots + A^{p^{e(fk-1)}} \equiv 0 \pmod P$$

furnishes a necessary and sufficient condition that (11.14) *be solvable in polynomials with coefficients in* $GF(p^n)$.

But by (11.07), the left member of (11.15)

$$(11.16) \qquad \begin{aligned} &\equiv \rho(A) + \rho^{p^e}(A) + \cdots + \rho^{p^{e(f-1)}}(A) \\ &\equiv c + c^{p^e} + \cdots + c^{p^{e(f-1)}}, \end{aligned}$$

c as in Theorem 11.4.

Now for all c in $GF(p^n)$, (11.16) is *equal* (not merely congruent) to c', a quantity in $GF(p^e)$. To evaluate c', suppose that $R(y)$ is an irreducible polynomial of degree f with coefficients in $GF(p^e)$. Then $R(y) = 0$ defines $GF(p^n)$. Write $c = c(y)$, a polynomial in y with coefficients in $GF(p^e)$. Theorem 11.4 may be applied, and we have

THEOREM 11.5. *If* $c = c(y)$ *is defined by* (11.12), *and*

$$c(y)R'(y) \equiv c'y^{f-1} + c_1'y^{f-2} + \cdots + c_{f-1}' \pmod{R(y)},$$

then (11.16) *reduces to* c'. *A necessary and sufficient condition that* (11.14) *be solvable is furnished by* $c' = 0$.

DUKE UNIVERSITY.

ORTHOGONALITY IN LINEAR METRIC SPACES

By Garrett Birkhoff

1. **Statement of main theorem.** Let B be any linear metric space[1] of three dimensions, whose points we shall suppose mapped linearly onto those of ordinary space.

It is natural to call a vector $\overline{pq}$ issuing from a point p of B "perpendicular" to a second such vector $\overline{pr}$ [in symbols, $\overline{pq} \perp \overline{pr}$] if and only if there is no point on the extended line through $\overline{pr}$ nearer to q than p.

Remark. Since translations of space are isometric, and uniform expansions about the origin multiply all distances by a constant factor of proportionality, $\overline{pq} \perp \overline{pr}$ implies that any vector parallel or anti-parallel to $\overline{pq}$ is perpendicular to any vector issuing from the same point and parallel or antiparallel to $\overline{pr}$. Therefore it is legitimate to say that the *direction* of $\overline{pq}$ is perpendicular to the *direction* of $\overline{pr}$.

The main purpose of this paper is to prove

Theorem 1. *If $\overline{pq} \perp \overline{pr}$ implies $\overline{pr} \perp \overline{pq}$, and if there is at most one perpendicular from a given line to a point not on that line, then B is "equivalent" to cartesian space (i.e., isometric with it under a linear transformation).*

2. **Outline of proof.** The proof of Theorem 1 involves such simple ideas that it is sufficient to sketch it.

First, let us fix on a particular linear representation of B in ordinary space. It is clear that the metric of B is determined by the "unit pseudo-sphere" S of points whose absolute values (in the terminology of von Neumann) are unity. It is also clear that S is a convex surface.

The argument then proceeds in two main steps. First it is shown that relative to any choice of cylindrical coördinates, the equation defining S is of the form

$$r = f(z) \cdot g(\theta). \tag{1}$$

Then it is shown (in effect) that any plane section of such a surface is an ellipse, essentially completing the proof.

To establish equation (1), let us first note that the radius $\overline{os}$ from the origin o to any point s on S is perpendicular to every line in any plane of support of S at s. Hence by the uniqueness and reciprocity of perpendicularity, S can have at most one plane of support at s.

Received March 10, 1935.

[1] As defined for instance by J. von Neumann, *On complete topological spaces*, Trans. Am. Math. Soc., vol. 37 (1935), pp. 3-4. The reader's attention is called to the definition of orthogonality in B. D. Roberts' *On the geometry of abstract vector spaces*, Tôhoku Math. Jour., vol. 39 (1934), pp. 42-59, which is essentially different.

Consequently by an elementary (and furthermore obvious) lemma[2] on convex surfaces, S is a smooth convex surface having a unique continuously varying tangent plane, whose direction maps S onto the sphere continuously. It follows incidentally that the perpendiculars to any given direction form a plane, and that all the perpendiculars in any given plane are perpendicular to a suitable line—which is unique by assumption.

It is hence possible to pick the z-axis in an entirely arbitrary direction in B, and then to pick x- and y-axes such that all three are mutually perpendicular.

Let us suppose that this has been done, and that in addition cylindrical coördinates have been set up, satisfying

$$(2) \qquad x = r\cos\theta, \qquad y = r\sin\theta, \qquad z = z.$$

Consider the plane curves of intersection γ_z of S with the planes $z = \text{const.}$ Because S is a smooth convex surface, they must be smooth convex curves. Besides, the tangential direction $T(z, \theta_0)$ to γ_z at $\theta = \theta_0$ must be perpendicular to (A) the radius to γ_z at $\theta = \theta_0$, and (B) the z-axis [since it lies in the (x, y)-plane]. Therefore the plane of directions perpendicular to $T(z, \theta_0)$ must be the plane $\theta = \theta_0$, and independent of z.

It follows[3] that the curves γ_z are all similar, and that the equation for S may be written in the form (1). The rest of the argument is based exclusively on this fact.

First one considers the "great pseudo-circles" β_θ of intersection of S with the planes $\theta = \text{const.}$ By (1) any one of them can be rotated into any other under a linear transformation

$$(3) \qquad z' = z, \qquad \theta' = \theta + \text{const.}, \qquad r' = g(\theta')/g(\theta)\cdot r$$

between the planes on which they lie leaving the z-axis fixed.

Then one recalls that the choice of the z-axis (of rotation) is arbitrary. Therefore by iterated use of the construction just described, one can rotate the (x, z)-plane P first about the z-axis into the (y, z)-plane, then about the y-axis into the plane through the y-axis and any line $\alpha_\lambda : x = \lambda z,\ y = 0$ in the (x, z)-plane, and finally about α_λ back into P. Moreover, the following statements are true.

(a) All three transformations are linear, and hence their product is also.
(b) The product preserves orientation in P.
(c) The intersection β_0 of S with P goes into itself.
(d) The intersection of S with the z-axis goes into the intersection of S with α_λ.

The last two paragraphs show that there exists an orientation-preserving homogeneous linear transformation of P into itself leaving β_0 invariant and carrying the z-axis into any straight line $x = \lambda z$. But on the other hand, since

[2] Proved for instance in T. Bonnesen and W. Fenchel, *Theorie der konvexen Körper*, Berlin, 1934, p. 13, bottom.

[3] By integrating a homogeneous linear differential equation of the form $dr/d\theta = T(\theta)$, where $T(\theta)$ is continuous.

β_0 is compact and does not contain the origin, the group G of *all* orientation-preserving homogeneous linear transformations of P into itself leaving β_0 invariant, is bounded—and therefore[4] equivalent after a homogeneous linear transformation h to a subgroup S of the orthogonal group. But since S is transitive on the lines through the origin, it must be the entire orthogonal group—that is, the entire group of rigid rotations.

Consequently the set of the transforms under S of any given point p of β_0 is a circle, and, $h(\beta_0)$ being a simple closed curve containing this set, $h(\beta_0)$ is a circle. Therefore under a suitable linear representation of B in coördinate space, the "absolute value" of a point $(x, 0, z)$ is $(x^2 + z^2)^{1/2}$, and the pythagorean theorem is true.

But perpendicularity and absolute value in B are intrinsic; hence the pythagorean theorem is true under all representations.

The conclusion of Theorem 1 now follows by decomposing (i) any vector into its y-component and perpendicular (x, z)-component; (ii) the (x, z)-component into perpendicular x- and z-components.

3. **Applications.** We can apply Theorem 1 immediately to the classification of complete linear metric spaces ["B-spaces" of Banach] by proving

THEOREM 2. *Any complete linear metric space L of three or more dimensions, in which perpendicularity is reciprocal and unique, is characterized by its (finite or transfinite) cardinal "dimension-number" n. Its elements correspond to the sets $(x_1, \cdots, x_n)$ of n real numbers satisfying $\sum_i x_i^2 < +\infty$, in such a way that if $x = (x_1, \cdots, x_n)$ and $y = (y_1, \cdots, y_n)$ are any two elements of L, and λ is any real number, then*

(a) *$x + y$ is the element $(x_1 + y_1, \cdots, x_n + y_n)$;*

(b) *λx is the element $(\lambda x_1, \cdots, \lambda x_n)$;*

(c) *the distance between x and y is $[\sum_i (x_i - y_i)^2]^{1/2}$.*

To prove Theorem 2, it is sufficient to observe that by Theorem 1 we can define "inner products" having the properties usually attached to inner products in Hilbert space. For these involve at most three independent elements[5]—and so can be proved by use of Theorem 1 within a 3-space containing them. Once this has been done, the rest of the argument is familiar.[6]

Another application yields an interesting geometrical definition of differential geometries. If one defines "pseudo-differential geometries" as varieties in which distance is given by integrating a differential form

$$\int \xi(x_1, \cdots, x_n; dx_1/dt, \cdots, dx_n/dt)dt = \int ds \tag{4}$$

[4] J. von Neumann, *Almost periodic functions in a group*, Trans. Amer. Math. Soc., vol. 36 (1934), p. 465, Theorem 19.

[5] M. H. Stone, *Linear transformations in Hilbert space*, New York, 1932, pp. 3–4.

[6] Cf. the mimeographed lecture notes of J. von Neumann on *Operator theory*, Theorem 12.27, Princeton, 1934.

such that for fixed x, $\xi(x; dx/dt)$ behaves like an "absolute value" for dx/dt, a condition which amounts to requiring that the coördinates be non-singular and that smooth geodesics exist, then Theorem 2 can be restated as

THEOREM 3. *A pseudo-differential geometry is a differential geometry [that is, for fixed x, $\xi(x; dx/dt)$ is the square root of a positive definite quadratic function of the dx_i/dt] if and only if local perpendicularity is unique and reciprocal.*

If in a Finsler space F transversality (= orthogonality) is unique and reciprocal, it is easy to show that $\xi(x; dx/dt) = \xi(x; -dx/dt)$, so that F is a "pseudo-differential" geometry. Hence a Finsler space is a differential geometry if and only if transversality is unique and reciprocal.

A final application yields a system of postulates for the geometry of ordinary space, more geometrical than the purely arithmetical definition of cartesian geometry, and simpler than the approach of Euclid.

4. **Necessity of three dimensions.** It is interesting that Theorem 1 is definitely false without the restriction that B have three dimensions. In fact,

THEOREM 4. *There exist infinitely many intrinsically different metric linear spaces of two dimensions in which perpendicularity is reciprocal and unique.*

Without going into detail, one may observe that a smooth convex simple closed curve Ω in coördinate space is the image of the "unit pseudo-circle" of elements of absolute value unity of a metric linear plane, if and only if the diameter drawn through the origin parallel to the tangent at any point p of Ω cuts Ω in points where the tangents to Ω are parallel to the diameter through p. This is a property related to conjugacy in the diameters of an ellipse.

But *any* smooth convex arc Ω defined over the first quadrant, symmetrical about the half-ray $\theta = \pi/4$, and perpendicular to the x-axis at its intersection with it, can be uniquely continued by this very condition through the other three quadrants. And by the symmetry about $\theta = \pi/4$, this will lead to a simple closed curve carried into itself under reflection through the origin, and hence into a satisfactory "unit pseudo-circle".

SOCIETY OF FELLOWS, HARVARD UNIVERSITY.

NEW IDENTITIES IN CONFORMAL GEOMETRY

By J. Levine

Complete sets of identities have been obtained for the components of the affine and metric normal tensors and also for the components of the affine, metric, and projective curvature tensors. In addition to these identities, a complete set is known for the components of the first covariant derivative of the affine curvature tensor.[1]

In this paper, complete sets of identities are obtained for the components ${}^0B^i_{jkl}$ $(i, j, k, l = 0, 1, \cdots, n, \infty)$ of the complete conformal curvature tensor[2] with the exception of the components ${}^0B^0_{IJ\infty}$ and ${}^0B^I_{\infty J\infty}$ $(I, J = 1, \cdots, n)$. A complete set of identities including the components ${}^0B^0_{IJ\infty}$ and ${}^0B^I_{\infty J\infty}$ has not been written down explicitly because of the excessive complexity of the calculations required. A method of obtaining such identities has been indicated, however, at the end of section 4.

1. Starting with the components[3] G_{IJ} of the fundamental conformal tensor defined by

$$G_{IJ} = \frac{g_{IJ}}{|g_{IJ}|^{\frac{1}{n}}},$$

the components K^A_{BC} are defined by[4]

$$K^A_{BC} = \frac{1}{2}G^{AS}\left(\frac{\partial G_{BS}}{\partial x^C} + \frac{\partial G_{SC}}{\partial x^B} - \frac{\partial G_{BC}}{\partial x^S}\right), \tag{1.1}$$

and the K^A_{BC} have the transformation equations[5]

$${}^0C^I_{AB} = \left(K^S_{MN}\frac{\partial x^M}{\partial y^A}\frac{\partial x^N}{\partial y^B} + \frac{\partial^2 x^S}{\partial y^A \partial y^B}\right)\frac{\partial y^I}{\partial x^S} - \frac{1}{n}(\bar{\psi}_A\delta^I_B + \bar{\psi}_B\delta^I_A - \bar{G}_{AB}\bar{G}^{SI}\bar{\psi}_S) \tag{1.2}$$

under the transformation of coördinates

$$x^A = f^A(y^1, \cdots, y^n). \tag{1.3}$$

Received December 5, 1934.

[1] T. Y. Thomas, *Differential Invariants of Generalized Spaces*, Cambridge University Press, 1934, pp. 81, 114, 132, 138. This reference will be called T.

[2] T., p. 81.

[3] In this section and the next three, indices have the following ranges: small Latin 0, 1, 2, $\cdots$, n, ∞; capital Latin 1, 2, $\cdots$, n; Greek 0, 1, 2, $\cdots$, n.

[4] These components K^A_{BC} were originally found in a different form by J. M. Thomas, Proc. Nat. Acad. Sci., vol. 11 (1925), pp. 257–9. They were obtained in the form shown by T. Y. Thomas as indicated in T., p. 67.

[5] T., p. 68.

The $\bar{\psi}_A$ are defined by

$$\bar{\psi}_A = \frac{\partial \log (xy)}{\partial y^A} = \frac{\partial^2 x^B}{\partial y^A \partial y^S} \frac{\partial y^S}{\partial x^B}. \tag{1.4}$$

It is natural to attempt to choose the functions f^A so that the coördinates y so determined will have the characteristic property of normal coördinates, i.e., so that

$$^0C^I_{AB} y^A y^B = 0. \tag{1.5}$$

In this attempt, of course, we should form the left member of (1.5) from that of (1.2) and then try to determine a coördinate transformation (1.3) from the resulting differential equations, some suitable initial conditions being assumed. If we take for these initial conditions

$$x^I = q^I, \qquad y^I = 0, \qquad \frac{\partial x^I}{\partial y^J} = \delta^I_J, \qquad \bar{\psi}_S = 0, \tag{1.6}$$

it is possible to determine the coefficients

$$\left(\frac{\partial^s x^I}{\partial y^{A_1} \cdots \partial y^{A_s}}\right)_{y=0}, \qquad (s = 2, 3, \cdots). \tag{1.7}$$

For our purpose, it will be sufficient to consider the coördinates y defined implicitly by

$$x^I = q^I + y^I - \frac{1}{2!}(K^I_{AB})_q y^A y^B + \cdots + \frac{1}{r!}\left(\frac{\partial^r x^I}{\partial y^{A_1} \cdots \partial y^{A_r}}\right)_0 y^{A_1} \cdots y^{A_r}, \tag{1.8}$$

the coefficients of the y terms being (1.7). For $s = 2$ in (1.7), the corresponding coefficient has the value shown in (1.8). The coördinates y thus obtained from (1.8) will be called conformal normal coördinates of order r.[6] In normal coördinates of order r the left member of (1.5) will not equal zero but its expansion in powers of y will begin with powers of y^I at least as great as $r + 1$. We shall indicate this by writing[7]

$$^0C^I_{AB} y^A y^B = O(y^{r+1}). \tag{1.9}$$

In conformal normal coördinates the K^I_{AB} will be denoted by $^0C^I_{AB}$. Also we define quantities D by

$$D^I_{JKL_1 \cdots L_s} = \frac{\partial^s \, {}^0C^I_{JK}(0)}{\partial y^{L_1} \cdots \partial y^{L_s}}, \qquad (s = 1, 2, \cdots). \tag{1.10}$$

[6] For a more detailed discussion of these coördinates, see the dissertation of V. A. Hoyle, *Some problems in conformal geometry*, Princeton University Library.

[7] The symbol $O(y^k)$ will be used to mean a series in the y^I beginning with powers of y^I at least as great as k.

It is found that the components ${}^0\Gamma^i_{j\alpha}$ of the associated conformal connection[8] when evaluated at the origin of normal coördinates[9] have the values

$$(1.11)\qquad \begin{gathered} {}^0C^i_{k0}(0) = -\frac{1}{n}\delta^i_k, \qquad {}^0C^i_{0\gamma}(0) = -\frac{1}{n}\delta^i_\gamma, \qquad {}^0C^\infty_{AB}(0) = -\frac{1}{n}G_{AB}, \\ {}^0C^I_{AB}(0) = {}^0C^0_{\infty A}(0) = {}^0C^\infty_{\infty A}(0) = {}^0C^0_{AB}(0) = {}^0C^A_{\infty B}(0) = 0. \end{gathered}$$

If we indicate by ${}^0G_{IJ}$ the component G_{IJ} in normal coördinates, (1.9) for $r = 3$ may be written by using (1.1)

$$(1.12)\qquad \left(2\frac{\partial\, {}^0G_{IJ}}{\partial y^K} - \frac{\partial\, {}^0G_{JK}}{\partial y^I}\right) y^J y^K = O(y^4).$$

The equations (1.12) characterize the y's as conformal normal coördinates of order 3. The ${}^0G$'s must of course satisfy the relations

$$(1.13)\qquad {}^0G_{IJ} = {}^0G_{JI},$$

$$(1.14)\qquad |{}^0G_{IJ}| = 1,$$

these conditions holding in all coördinate systems. The relations (1.13) and (1.14) constitute a complete set of identities for the components ${}^0G_{IJ}$. Now the ${}^0G_{IJ}$ have the power series expansions

$$(1.15)\qquad {}^0G_{IJ} = {}^0G_{IJ}(0) + \frac{1}{2!}\,{}^0G_{IJ,KL}(0) y^K y^L + \cdots.$$

If we form the derivatives $\partial^0 G_{IJ}/\partial y^K$ from (1.15) and substitute them in (1.12), it will be found that we must have

$$(1.16)\qquad 2(G_{IJ,KL} + G_{IL,JK} + G_{IK,JL}) - (G_{JK,IL} + G_{KL,IJ} + G_{JL,IK}) = 0.$$

The quantities $G_{IJ,KL}$ are defined by

$$G_{IJ,KL} = \frac{\partial^2\, {}^0G_{IJ}(0)}{\partial y^K \partial y^L}.$$

If in (1.16) the indices I, K and J, L be interchanged and the resulting expression be subtracted from (1.16), there results

$$(1.17)\qquad G_{IJ,KL} - G_{KL,IJ} + G_{IL,JK} - G_{JK,IL} = 0.$$

On interchanging I, J and K, L in (1.17) and subtracting we then obtain

$$(1.18)\qquad G_{IJ,KL} = G_{KL,IJ},$$

[8] T., p. 70.

[9] In this section and the next the term normal coördinates will mean conformal normal coördinates of order 3. In sections 3 and 4 it will mean conformal normal coördinates of order 4.

and when use of (1.18) is made in (1.16) we get

(1.19) $$G_{IJ,KL} + G_{IK,LJ} + G_{IL,JK} = 0 .$$

Conversely, from (1.19) follows (1.16), and also (1.18). The $G_{IJ,KL}$ also satisfy the symmetry identities

(1.20) $$G_{IJ,KL} = G_{JI,KL} ,$$

(1.21) $$G_{IJ,KL} = G_{IJ,LK} .$$

These identities were used in obtaining (1.19) from (1.16).

On differentiating (1.14) twice and then evaluating at the origin of the normal coördinates, it is found that

(1.22) $$G^{AB}\, G_{AB,CD} = 0 ,$$

use being made of

(1.23) $$G_{IJ,K} \equiv \left(\frac{\partial\, {}^0G_{IJ}(0)}{\partial y^K}\right) = 0 ,$$

this latter relation resulting from (1.11) and (1.1), i.e.,

(1.24) $$\frac{\partial\, {}^0G_{IJ}}{\partial y^K} = {}^0G_{IS}\, {}^0C^S_{JK} + {}^0G_{SJ}\, {}^0C^S_{IK} .$$

It will be shown that (1.19), (1.20), (1.21) and (1.22) together with (1.13), (1.14) with ${}^0G_{IJ}$ replaced by G_{IJ} constitute a complete set of identities for the components G_{IJ} and $G_{IJ,KL}$. Consider then the sets of numbers

$$G_{IJ},\ G_{IJ,KL}$$

chosen so as to satisfy the above mentioned identities and furthermore such that the series

(1.25) $${}^0G_{IJ} = G_{IJ} + \frac{1}{2!}\, G_{IJ,KL}\, y^K y^L + \cdots$$

converge, and also so that $|\, {}^0G_{IJ} \,| = 1$. Then the ${}^0G_{IJ}$ so defined satisfy the conditions (1.12) which characterize the y^I as a set of conformal normal coördinates of order 3. This completes the proof.

2. By evaluating the components ${}^0B^I_{ABC}$ of the complete conformal curvature tensor with components ${}^0B^i_{abc}$ at the origin of normal coördinates and using (1.10) it is found that

(2.1) $${}^0B^I_{ABC} = D^I_{ABC} - D^I_{ACB} .$$

From the identities

(2.2) $$D^I_{ABC} = D^I_{BAC} ,$$

2.3) $$D^I_{ABC} + D^I_{BCA} + D^I_{CAB} = 0 ,$$

which are obtained by use of (1.9) and (1.10), we are able to obtain from (2.1)

$$D^I_{ABC} = \frac{1}{3}\left({}^0B^I_{ABC} + {}^0B^I_{BAC}\right). \tag{2.4}$$

Since the components ${}^0B^I_{ABC}$ are those of a tensor (the Weyl conformal curvature tensor), (2.4) shows the components D^I_{ABC} are also those of a tensor. The higher order D's, however, do not have this property in general. From (1.24) by differentiation there is obtained

$$G_{AB,CD} = G_{AE}\, D^E_{BCD} + G_{EB}\, D^E_{ACD}\,, \tag{2.5}$$

and by substituting for the D's in (2.5) from (2.4) we get

$$G_{AB,CD} = \frac{1}{3}\,({}^0B_{ACBD} + {}^0B_{BCAD})\,, \tag{2.6}$$

where

$${}^0B_{ABCD} = G_{AI}\, {}^0B^I_{BCD}\,. \tag{2.7}$$

By expressing (1.1) in normal coördinates, differentiating and evaluating at the origin we obtain, on using (1.10) and

$$\left(\frac{\partial\, {}^0G^{IJ}(0)}{\partial y^K}\right) = 0\,, \tag{2.8}$$

$$D^I_{JKL} = \frac{1}{2} G^{IA}(G_{AJ,KL} + G_{AK,JL} - G_{JK,AL})\,. \tag{2.9}$$

Hence, from (2.1) and (2.7) we get

$${}^0B_{ABCD} = G_{AC,BD} - G_{AD,BC}\,. \tag{2.10}$$

If in (2.10) we interchange the indices A, B and add, there results with the aid of (1.18)

$${}^0B_{ABCD} + {}^0B_{BACD} = 0\,. \tag{2.11}$$

Also, by interchanging C and D and adding we obtain

$${}^0B_{ABCD} + {}^0B_{ABDC} = 0\,. \tag{2.12}$$

With the use of (1.20) and (1.21) we next obtain the identities

$${}^0B_{ABCD} + {}^0B_{ACDB} + {}^0B_{ADBC} = 0\,. \tag{2.13}$$

Finally we form $G^{BC}\, {}^0B_{ABCD}$, and from (2.10) and (1.22) get

$$G^{BC}\, {}^0B_{ABCD} = G^{BC}\, G_{AB,CD}\,,$$

but from (1.19) it is easily shown that

$$G^{BC}\, G_{AB,CD} = 0\,,$$

thus giving the identities

$$G^{BC}\,{}^0B_{ABCD} = 0. \tag{2.14}$$

The relations (2.6) can also be obtained from (2.10) and the identities on the components $G_{AB,CD}$. Thus from (2.10) and (1.19), (1.20), (1.21), (1.22) we can obtain (2.6) and (2.11), (2.12), (2.13), (2.14). Conversely, it is not difficult to show that the first group of the above relations can be obtained from the second. From this we have

[2.1] *The identities* (2.11), (2.12), (2.13), (2.14) *and* $G_{IJ} = G_{JI}$, $| G_{IJ} | = 1$ *constitute a complete set of identities for the components* G_{IJ}, ${}^0B_{ABCD}$.

From the identities for ${}^0B_{ABCD}$, the following may be derived

$$ {}^0B_{ABCD} = {}^0B_{CDAB}\,. \tag{2.15}$$

It is interesting to note that the components of the ordinary affine or metric curvature tensor satisfy identities of the types (2.11), (2.12), (2.13).

From their identities, it can be shown that the number of independent components ${}^0B_{ABCD}$ is given by

$$\frac{1}{12}n(n+1)(n+2)(n-3)\,.$$

This number vanishes for $n = 3$, a fact which agrees with the well-known result that the Weyl conformal curvature tensor is zero for $n = 3$.

3. In this section the complete set of identities for the quantities $G_{IJ,\,KLM}$ will be found, and then from them the identities of the remaining components ${}^0B^{\,i}_{jkl}$ to be considered will be obtained. The quantities $G_{IJ,\,KLM}$ are defined by

$$G_{IJ,\,KLM} = \frac{\partial^3\,{}^0G_{IJ}(0)}{\partial y^K\,\partial y^L\,\partial y^M}\,.$$

By using conformal normal coördinates of order four and proceeding in a manner similar to that used for the $G_{IJ,\,KL}$ in section one, it will be found that the components $G_{AB,\,CDE}$ satisfy the identities

$$\begin{aligned} &3(G_{AB,\,CDE} + G_{AC,\,BDE} + G_{AD,\,BCE} + G_{AE,\,BCD}) \\ &\quad - (G_{BC,\,DEA} + G_{BD,\,CEA} + G_{BE,\,CDA} + G_{CD,\,BEA} + G_{CE,\,BDA} + G_{DE,\,BCA}) = 0. \end{aligned} \tag{3.1}$$

By replacing the indices $ABCDE$ in (3.1) by the sets $BCDEA$, $CDEAB$, $DEABC$ in turn and adding the resulting three sets of equations to (3.1), we obtain

$$G_{AB,\,CDE} + G_{AC,\,DEB} + G_{AD,\,EBC} + G_{AE,\,BCD} = 0\,. \tag{3.2}$$

Use is also made of the symmetry identities

$$G_{AB,\,CDE} = G_{BA,\,CDE}\,, \qquad G_{AB,\,CDE} = G_{AB,\,PQR}\,, \tag{3.3}$$

where PQR represents any permutation of CDE. By using (3.2) and (3.1) we then obtain

(3.4) $$G_{AB,CDE} + G_{AC,BDE} + G_{AD,BCE} + G_{BC,ADE} + G_{BD,ACE} + G_{CD,ABE} = 0.$$

From (1.13), (1.14) we find

(3.5) $$G^{AB}G_{AB,CDE} = 0.$$

It can be shown, in a manner analogous to that used for the components $G_{IJ,KL}$ at the end of section two, that (3.2), (3.3), (3.5) and $G_{IJ} = G_{JI}$, $|G_{IJ}| = 1$ constitute a complete set of identities for the components G_{IJ} and $G_{AB,CDE}$. From (3.4) it is found that the following identities are satisfied

(3.6) $$G^{AC}G^{BD}G_{AB,CDE} = 0.$$

By differentiating (1.24) and evaluating at the origin of normal coördinates of order four, there is obtained

(3.7) $$G_{AB,CDE} = G_{AM}D^{M}_{BCDE} + G_{MB}D^{M}_{ACDE}.$$

Also, from (1.1) in normal coördinates we get by differentiating twice and evaluating at the origin

(3.8) $$D^{I}_{JKLM} = \tfrac{1}{2}G^{IS}(G_{SJ,KLM} + G_{SK,LMJ} - G_{JK,LMS}).$$

On placing

$$D_{ABCDE} = G_{IE}D^{I}_{ABCD},$$

it can be shown, by use of (3.8), that we have the identities

(3.9) $$D_{ABCDE} = D_{BACDE} = D_{ABDCE}; \qquad D_{ABCDE} + D_{EBCDA} = D_{ACBDE} + D_{ECBDA},$$

(3.10) $$G^{AE}D_{ABCDE} = 0,$$

(3.11) $$D_{ABCDE} + D_{ACBDE} + D_{ADBCE} + D_{BCADE} + D_{BDACE} + D_{CDABE} = 0.$$

By expressing ${}^{0}B^{i}_{k\alpha\beta}$ in normal coördinates of order three we obtain[10]

(3.12) $${}^{0}V^{i}_{k\alpha\beta} = \frac{\partial\, {}^{0}C^{i}_{k\alpha}}{\partial y^{\beta}} - \frac{\partial\, {}^{0}C^{i}_{k\beta}}{\partial y^{\alpha}} + {}^{0}C^{i}_{j\beta}\,{}^{0}C^{j}_{k\alpha} - {}^{0}C^{i}_{j\alpha}\,{}^{0}C^{j}_{k\beta}.$$

In (3.12), ${}^{0}V$ represents ${}^{0}B$ in the y coördinates and ${}^{0}C^{i}_{k\alpha}$ the components of the associated conformal connection. If now (3.12) be evaluated at the origin of the normal coördinates, we obtain

(3.13) $${}^{0}B^{i}_{k00} = {}^{0}B^{i}_{kA0} = {}^{0}B^{0}_{\infty AB} = {}^{0}B^{\alpha}_{0AB} = {}^{0}B^{\infty}_{iAB} = 0,$$

(3.14) $${}^{0}B^{I}_{ABC} = D^{I}_{ABC} - D^{I}_{ACB},$$

(3.15) $${}^{0}B^{0}_{ABC} = \frac{n}{n-2}(D^{S}_{ABCS} - D^{S}_{ACBS}),$$

[10] T., p. 72.

$$(3.16) \qquad {}^0B^I_{\infty BC} = \frac{n}{n-2} G^{IT}(D^S_{TBCS} - D^S_{TCBS}),$$

$$(3.17) \qquad {}^0B^i_{k\alpha\beta} = -{}^0B^i_{k\beta\alpha}.$$

Thus, the only non-zero components of ${}^0B^i_{k\alpha\beta}$ are

$${}^0B^I_{ABC}, \qquad {}^0B^0_{ABC}, \qquad {}^0B^I_{\infty BC},$$

and from (3.15), (3.16) we have

$$(3.18) \qquad {}^0B^I_{\infty BC} = G^{IA}\,{}^0B^0_{ABC}.$$

To find the remaining ${}^0B^i_{abc}$ other than ${}^0B^i_{a\beta\gamma}$, we use the definitions[11]

$$(3.19) \quad {}^0B^i_{kA\infty} = -{}^0B^i_{k\infty A}; \qquad (3.19\text{a}) \quad {}^0B^i_{k0\infty} = {}^0B^i_{k\infty 0} = {}^0B^i_{k\infty\infty} = 0,$$

$$(3.20) \qquad {}^0B^i_{kA\infty} = \frac{n}{4-n}\,{}^0B^i_{kA,EF}G^{EF} \qquad (n \neq 4),$$

where

$$(3.21) \quad {}^0B^i_{kA,EF} = \frac{\partial\,{}^0B^i_{kAE}}{\partial x^F} + {}^0B^h_{kAE}\,{}^0\Gamma^i_{hF} - {}^0B^i_{hAE}\,{}^0\Gamma^h_{kF} - {}^0B^i_{k\sigma E}\,{}^0\Gamma^\sigma_{AF} - {}^0B^i_{kA\sigma}\,{}^0\Gamma^\sigma_{EF}.$$

In what follows, it will be assumed $n \neq 4$ as ${}^0B^i_{kA\infty}$ is not defined for $n = 4$. From (3.21) we calculate

$$(3.22) \qquad \begin{aligned} &{}^0B^0_{0A,EF} = \frac{1}{n}\,{}^0B^0_{FAE}, \quad {}^0B^\infty_{0A,EF} = {}^0B^0_{\infty A,EF} = 0, \quad {}^0B^\infty_{\infty A,EF} = -\frac{1}{n}G_{HF}\,{}^0B^H_{\infty AE}, \\ &{}^0B^I_{0A,EF} = \frac{1}{n}\,{}^0B^I_{FAE}, \qquad {}^0B^\infty_{KA,EF} = -\frac{1}{n}G_{HF}\,{}^0B^H_{KAE}. \end{aligned}$$

With the aid of (2.14), (3.17), (3.20), (3.22), and

$$(3.23) \qquad G^{AB}\,{}^0B^0_{ABC} = 0,$$

$$(3.24) \qquad {}^0B^I_{\infty IA} = 0,$$

the proofs of which will be given later, it follows that

$$(3.25) \qquad {}^0B^i_{0A\infty} = {}^0B^\infty_{AB\infty} = {}^0B^\infty_{\infty A\infty} = {}^0B^0_{\infty A\infty} = 0.$$

From the results so far obtained it can be seen that the only non-vanishing components ${}^0B^i_{abc}$ are

$${}^0B^\alpha_{IJK}, \qquad {}^0B^\alpha_{IJ\infty}, \qquad {}^0B^I_{\infty JK}, \qquad {}^0B^I_{\infty J\infty}.$$

The components ${}^0B^A_{IJK}$ have already been discussed. We shall now consider the components ${}^0B^0_{IJK}$, ${}^0B^A_{IJ\infty}$, ${}^0B^I_{\infty JK}$.

[11] T., p. 77.

4. By means of (3.15) and (3.8) we find that

$$(4.1) \qquad {}^0B^0_{IJK} = \frac{n}{2(n-2)} G^{ST}(G_{JS,IKT} + G_{IK,JST} - G_{IJ,KST} - G_{KS,IJT}).$$

Instead of using ${}^0B^I_{JK\infty}$, we shall use its covariant form

$${}^0B_{IJK\infty} = G_{IS}\,{}^0B^S_{JK\infty}.$$

From (3.20), (3.21) we obtain after some calculations

$$(4.2) \quad {}^0B_{LKA\infty} = \frac{n}{4-n}\left[G^{EF}(D_{KAEFL} - D_{KEFAL}) + \frac{2}{n-2}(D^S_{ALKS} - D^S_{AKLS})\right].$$

In obtaining (4.2) we use the relations

$$G^{EF}D^S_{AEFS} = 0,$$

which follow from (3.8) and (3.6). Now if in (4.2) we replace the D's by means of (3.8) we find after reducing and using (4.1) that

$$(4.3) \qquad {}^0B_{IJK\infty} = {}^0B^0_{KIJ}.$$

By means of (3.18) we can express the ${}^0B^I_{\infty BC}$ in terms of the G_{IJ} and ${}^0B^0_{IJK}$. Thus all the components besides ${}^0B^0_{IJK}$ under consideration, i.e., ${}^0B^I_{\infty JK}$, ${}^0B^0_{IJK}$ can be expressed in terms of the G_{IJ} and ${}^0B^0_{IJK}$. In order to solve for the $G_{IJ,KLM}$ in terms of the ${}^0B$, we introduce the covariant derivatives ${}^0B_{ABCD,E}$ of ${}^0B_{ABCD}$. These are given by[12]

$$(4.4) \quad \begin{aligned} {}^0B_{IJKL,M} = G_{IS}\Bigg[\frac{\partial\,{}^0B^S_{JKL}}{\partial x^M} &+ {}^0B^h_{JKL}\,{}^0\Gamma^S_{hM} - {}^0B^S_{hKL}\,{}^0\Gamma^h_{JM} \\ &- {}^0B^S_{JhL}\,{}^0\Gamma^h_{KM} - {}^0B^S_{JKh}\,{}^0\Gamma^h_{LM}\Bigg]. \end{aligned}$$

By expressing (4.4) in normal coördinates and then evaluating at the origin, we obtain by the use of (1.11), (3.12), (3.18), (4.3)

$$(4.5) \qquad D_{JKLMI} - D_{JLKMI} = \phi_{JKLMI},$$

where

$$(4.6) \quad \begin{aligned} \phi_{JKLMI} = {}^0B_{IJKL,M} &+ \frac{1}{3n}[G_{IL}({}^0B^0_{JKM} + {}^0B^0_{KJM}) + G_{JK}({}^0B^0_{ILM} + {}^0B^0_{LIM}) \\ &- G_{KI}({}^0B^0_{JLM} + {}^0B^0_{LJM}) - G_{JL}({}^0B^0_{IKM} + {}^0B^0_{KIM})] \\ &+ \frac{1}{n}[G_{IM}\,{}^0B^0_{JKL} + G_{KM}\,{}^0B^0_{LIJ} - G_{JM}\,{}^0B^0_{IKL} - G_{LM}\,{}^0B^0_{KIJ}]. \end{aligned}$$

[12] T., p. 74.

We now use (4.5) in conjunction with (3.9), (3.11) and obtain

$$(4.7)\qquad 6D_{JKLMI} = 2\phi_{JKLMI} + \phi_{JKMLI} + \phi_{KJLMI} + \phi_{KJMLI} + \phi_{LJMKI}.$$

We can now obtain the $G_{IJ,KLM}$ in terms of the ${}^0B$ by substituting from (4.7) in (3.7). This gives

$$(4.8)\quad 6G_{IJ,KLM} = \phi_{KJLMI} + \phi_{KILMJ} + \phi_{KJMLI} + \phi_{KIMLJ} + \phi_{LJMKI} + \phi_{LIMKJ}.$$

To obtain identities for the ${}^0B^0_{IJK}$, we use the equations (4.1) and the identities for the $G_{IJ,KLM}$. We can get the following

$$(4.9)\qquad {}^0B^0_{IJK} = -{}^0B^0_{IKJ}, \qquad {}^0B^0_{IJK} + {}^0B^0_{JKI} + {}^0B^0_{KIJ} = 0, \qquad G^{IJ}\,{}^0B^0_{IJK} = 0.$$

This proves (3.23), and (3.24) then follows from (3.18). From (4.5) and (3.8) we now have

$$(4.10)\qquad \phi_{JKLMI} = \tfrac{1}{2}(G_{IK,JLM} + G_{JL,IKM} - G_{JK,LMI} - G_{IL,KMJ}).$$

Hence from (4.10), (4.1) and (4.6) we can express ${}^0B_{IJKL,M}$ in terms of the G's. We indicate this by

$$(4.11)\qquad {}^0B_{IJKL,M} = \phi_{JKLMI} - \theta_{JKLMI},$$

where the θ represents the remaining terms in the right member of (4.6) after excluding ${}^0B_{IJKL,M}$. The θ is to be thought of as expressed in terms of the G's and likewise for the ϕ. It is easily shown that the ϕ satisfy

$$(4.12)\qquad \begin{gathered} \phi_{JKLMI} = -\phi_{JLKMI} = -\phi_{IKLMJ}, \\ \phi_{JKLMI} + \phi_{JLMKI} + \phi_{JMKLI} = 0, \qquad \phi_{JKLMI} + \phi_{KLJMI} + \phi_{LJKMI} = 0. \end{gathered}$$

These have as a consequence

$$(4.13)\qquad \phi_{JKILM} = \phi_{IMJLK}.$$

The θ's also satisfy the same identities (4.12) as the ϕ's. Hence from (4.11) we get

$$(4.14)\qquad \begin{gathered} {}^0B_{IJKL,M} = -{}^0B_{IJLK,M}, \qquad {}^0B_{IJKL,M} = -{}^0B_{JIKL,M}, \\ {}^0B_{IJKL,M} + {}^0B_{IJLM,K} + {}^0B_{IJMK,L} = 0, \\ {}^0B_{IJKL,M} + {}^0B_{IKJL,M} + {}^0B_{ILJK,M} = 0. \end{gathered}$$

From (3.5) and (4.8) we get

$$(4.15)\qquad G^{IJ}(\phi_{KILMJ} + \phi_{KIMLJ} + \phi_{LJMKI}) = 0,$$

and replacing the ϕ's in (4.15) from (4.6) results in

$$(4.16)\qquad G^{IJ}({}^0B_{IKJL,M} + {}^0B_{IKJM,L} + {}^0B_{ILJM,K}) = 0.$$

One more identity will be needed, this being obtained as follows. From (4.1) and (4.10) we obtain

$$G^{ST}\phi_{ABCST} = \frac{n-2}{n}\,{}^0B^0_{ABC}\,. \tag{4.17}$$

Now if in (4.8) we replace the ϕ's by means of (4.6) and then form the expression

$$G^{ST}\phi_{ABCST}$$

from the left members of (4.8), we find after considerable reduction

$$G^{ST}\phi_{ABCST} = \frac{n+1}{n}\,{}^0B^0_{ABC} + G^{ST}\,{}^0B_{SABC,T}\,.$$

On comparing the above equation with (4.17), we obtain the desired identity

$$G^{ST}\,{}^0B_{SABC,T} = -\frac{3}{n}\,{}^0B^0_{ABC}\,. \tag{4.18}$$

From the identities (4.9), (4.14), (4.16), (4.18) it is possible to obtain the complete set of identities of the $G_{IJ,KLM}$, use being made of the defining relation (4.8), in which the ϕ's are to be replaced from (4.6). In obtaining the second symmetry identities of (3.3) it is sufficient to show

$$G_{AB,CDE} = G_{AB,DCE} = G_{AB,CED}\,,$$

for from these the remaining identities involving further permutations of the indices CDE follow. It can also be shown that from (4.8), (4.9), (4.14), (4.16), (4.18) the relations (4.1) and (4.11) can be obtained. We can now state, $(n \neq 4)$.

[4.1] *The identities* (2.11), (2.12), (2.13), (2.14), (3.13), (3.18), (3.19a), (3.25), (4.3), (4.9), (4.14), (4.16), (4.18) *and* $G_{IJ} = G_{JI}$, $|G_{IJ}| = 1$ *constitute a complete set of identities for the components* G_{IJ}, ${}^0B^i_{jkl}$, ${}^0B_{IJKL,M}$ *excluding the components* ${}^0B^0_{IJ\infty}$, ${}^0B^I_{\infty J\infty}$.

For all $n \geq 3$, the identities mentioned in [2.1] and (3.18), (1.9) are a complete set for the non-zero components of the incomplete conformal curvature tensor ${}^0B^i_{j\alpha\beta}$ and G_{IJ}.

With regard to the excluded components ${}^0B^0_{IJ\infty}$, ${}^0B^I_{\infty J\infty}$ it can be shown that the following relations are satisfied $(n \neq 4)$

$${}^0B^0_{AB\infty} = G_{AI}\,{}^0B^I_{\infty B\infty}\,, \tag{4.19}$$

$$G^{AB}\,{}^0B^0_{AB\infty} = {}^0B^I_{\infty I\infty} = 0, \tag{4.20}$$

$$\begin{aligned}\frac{(4-n)(n-2)}{n^2}\,{}^0B^0_{AB\infty} &= G^{CD}(D^S_{ABCDS} - D^S_{ACDBS}) \\ &+ \frac{1}{2(n-1)}\,G^{EF}(D^S_{EFBAS} - G_{AB}\,G^{CD}\,D^S_{EFCDS}) \\ &+ \frac{1}{n-1}\,G^{EF}(G_{AB}\,G^{CD}\,D^S_{TFD}\,D^T_{SEC} - D^S_{TFA}\,D^T_{SEB}) \\ &+ G^{CD}(D^T_{SAD}\,D^S_{TCB} - 2D^S_{TBD}\,D^T_{SAC}).\end{aligned}$$

From (1.1) in conformal normal coördinates y^I of order five, by differentiation we can obtain expressions for the components D^I_{ABCDE} in terms of the quantities

$$G_{IJ,\,KLMP} \equiv \frac{\partial^4\,{}^0G_{IJ}(0)}{\partial y^K\,\partial y^L\,\partial y^M\,\partial y^P},$$

and hence can express ${}^0B^0_{AB\infty}$ in terms of these quantities. The $G_{IJ,\,KLMP}$ can be expressed in the form

$$G_{IJ,\,KLMP} = H_{IJKLMP}(G_{AB};\,{}^0B^i_{jkl};\,{}^0B^i_{jkl,m};\,{}^0B^I_{JKL,M,P}). \tag{4.21}$$

Denote by (β) the components composing the H's. A complete set of identities for the components $G_{IJ,\,KLMP}$ can be obtained in a manner similar to that used for the $G_{IJ,KL}$ and $G_{IJ,KLM}$. Now by adding to the identities mentioned in [4.1] suitable identities between the components (β), we can obtain by means of the resulting set of identities and the defining relations (4.21) the complete set of identities for the $G_{IJ,\,KLMP}$ and also the relations defining the components (β) in terms of the G_{IJ}, $G_{IJ,\,KL}$, $G_{IJ,\,KLM}$, $G_{IJ,\,KLMP}$. The totality of identities in the components (β) thus used would constitute a complete set of identities for these components. As part of these, we should use (4.19) and (4.20).

PRINCETON UNIVERSITY.

METABELIAN GROUPS AND TRILINEAR FORMS

By H. R. Brahana

1. The relation of the theory of metabelian groups to the fundamental problem of finite groups. The theory of abstract groups is not yet one hundred years old. One hundred years is a short time in the history of mathematics, and yet the theory of groups is old compared with many subjects that are receiving the attention of large numbers of mathematicians. The greater part of the literature on finite groups is concerned with finding properties of certain known groups and with finding classes of groups which have certain given properties. This work, while necessary to the unfolding of the theory of groups, contributes usually only indirectly and often very remotely to the solution of the fundamental problem. The fundamental problem of finite groups is the determination of all the groups of a given order n. The determination of all the groups of a given order n will in general consist of the determination of several sets of properties such that each group of order n possesses all the properties of one set, and such that a group which possesses all the properties of one set does not possess all the properties of any other set. The determination of a definitive set of properties for a group does not close the subject of that particular group, for the reasons that there are in general large numbers of definitive sets for a given group and any given set may be entirely unsuitable for a particular purpose.

Although the fundamental problem was recognized and attacked at an early date, progress toward its solution has been very slow. In 1854 Cayley proved[1] that there are but two groups of order four and two groups of order six. In 1930 G. A. Miller listed the groups of orders up to 100 including those of order 64 which had not been completely determined before.[2] In 1934 Senior and Lunn[3] listed the groups from order 101 to 161 omitting those of order 128. Apparently nobody claims to have determined the groups of order p^7, and according to Miller's paper cited above Potron's attempt[4] to determine the groups of order p^6 was unsuccessful for $p = 2$. We do not wish to suggest that progress toward solution of the fundamental problem should be measured by that num-

Received February 21, 1935.

[1] *On the theory of groups as depending on the symbolical equation* $\theta^n = 1$, Philosophical Magazine, vol. 7 (1854), pp. 40–47; *Collected Mathematical Papers*, vol. II, pp. 123–130. For a sketch of the history of groups see the last paper in the forthcoming volume of the *Collected Works* of G. A. Miller, University of Illinois Press.

[2] *Determination of all the groups of order* 64, American Journal of Mathematics, vol. 52 (1930), pp. 617–634.

[3] J. K. Senior and A. C. Lunn, *Determination of the groups of orders* 101–161, *omitting order* 128, American Journal of Mathematics, vol. 56 (1934), pp. 328–338.

[4] M. Potron, *Les groupes d'ordre* p^6, Thesis, Paris, 1904.

ber n, yet on the other hand groups of order p^6 must be considered rather elementary among the groups of order p^n.

A given group brings with it a whole complex of other groups—subgroups, quotient groups, etc.—most of which are more elementary than the given group. If two groups are simply isomorphic the complex determined by one must be abstractly the same as that determined by the other. Comparison of the more elementary groups of the two complexes is the usual method of determining the non-isomorphism of two groups. Sylow's theorem narrows the region of attack on the fundamental problem by pointing out that progress for a general n must wait upon progress for the case where n is a power of a prime. Of the extensive literature on prime-power groups very little has been directed toward solution of difficulties that appeared early in connection with the fundamental problem.

The operators which are p^{th} powers in a group of order p^m generate a characteristic subgroup. The corresponding quotient group is characteristic and its operators are all of order p. If this quotient group is abelian, it is completely determined by its order; if it is non-abelian, it is said to be conformal with the abelian group of type $1, 1, \cdots$ and with any other such group of the same order. Thus the question of the simple isomorphism of two groups of prime-power order always involves the question of the isomorphism of two groups conformal with abelian groups of type $1, 1, \cdots$. The fundamental problem for groups of order p^m conformal with the abelian group of type $1, 1, \cdots$ entered the literature about 1900. For $p = 2$ there is no problem, since the only such group is the abelian group. Burnside touched upon the problem for $p = 3$,[5] although his paper has another purpose. He proved that a group generated by m operators of order 3 which contained only operators of order 3 was finite, and he stated a value for the order of the group provided the generators were independent except for the above conditions. This value was incorrect and the correct value was determined by Levi and van der Waerden in 1932.[6] Although the importance of the problem of groups containing only operators of order p was appreciated at least as early as Burnside's paper, nevertheless in 1933 Miller stated[7] (correctly) that the simplest case, for $p = 3$, had not yet been solved. These groups are all known to be of class 3. The

[5] W. Burnside, *On an unsettled question in the theory of discontinuous groups*, Quarterly Journal of Mathematics, vol. 33 (1902), pp. 230–238.

[6] F. Levi and B. L. van der Waerden, *Über eine besondere Klasse von Gruppen*, Hamburg Abhandlungen, vol. 9 (1932), pp. 154–158.

[7] *Groups involving a small number of squares*, Proceedings of the National Academy of Sciences, vol. 19 (1933), pp. 1054–1057. The 'correctly' in parentheses above is prompted by the review of Miller's paper in the Jahrbuch, vol. 59 (1933), p. 149. The reviewer implies that a solution of the problem is to be found in the paper by Levi and van der Waerden. Every group generated by m operators and containing only operators of order 3 is simply isomorphic with some quotient group of Levi and van der Waerden's group. This solves the problem about as adequately as it is solved by remarking that every group of order p^m is simply isomorphic with some subgroup of the symmetric group on p^m letters. If the problem is not to be ignored, it is necessary to determine the invariant subgroups of Levi and van der Waerden's group, and in the writer's opinion this is not feasible.

class of a group measures the series of successive central quotient groups. The groups of class 1 are abelian, those of class 2 are metabelian. Not only have the groups of class 3 not been determined, but neither have the groups of class 2. This fact seems to the writer to be the proper measuring-stick to use in evaluating the progress toward a solution of the fundamental problem.

Every group containing only operators of order p has all of its abelian subgroups of type $1, 1, \cdots$. Every such group contains an invariant abelian subgroup; for example, the central. Every invariant abelian subgroup is either maximal invariant abelian or is contained in a maximal invariant abelian subgroup. Every maximal invariant abelian subgroup contains the central. Let us denote the given group by G and a maximal invariant abelian subgroup by H. The quotient group G/H is simply isomorphic with some subgroup of I, the group of isomorphisms of H. The fact that two groups G and G' can be simply isomorphic only if G/H and G'/H' as well as H and H', are simply isomorphic shows that a study of the subgroups of order p^a, with operators of order p, in a Sylow subgroup I_p corresponding to the prime p in the group of isomorphisms of H is indispensable. This group I has probably had more attention than any other group of isomorphisms. It was noticed long ago that this group can be represented as the set of non-singular square matrices of n rows (H is of order p^n) with elements in $GF(p)$ and with a law of combination which is ordinary matrix multiplication.

Dickson observed[8] that the subset of these matrices with 1's on the main diagonal and 0's below constitutes a Sylow subgroup I_p of the group I. It is obvious that the characteristic polynomial of such a matrix is $(1 - \lambda)^n$. It has been known for a long time that the conjugacy of two such matrices in I is determined by the identity of their invariant factors. Striking as it may seem, these facts were never put together until 1932.[9] The result is a theorem which makes possible an attack on the fundamental problem. The theorem asserts that the conjugate set to which an operator of I_p belongs is characterized by a set of numbers $n_1, n_2, \cdots, n_s$, which are the degrees of the invariant factors of the corresponding matrix.

The simple isomorphism of G and G' above requires the simple isomorphism of G/H and G'/H' but it requires more. It requires that the subgroups of I_p determined by G/H and G'/H' be conjugate in I. The simple method of characterization of operators of I_p resulting from the theorem above furnishes a means of getting deeper into the question of conjugacy of subgroups. Two subgroups of I_p can be conjugate only if they contain operators of the same types. This is not a sufficient condition but it is a condition comparatively

[8] *Determination of all the subgroups of the three highest powers of p in the group G of all m-ary linear homogeneous transformations modulo p*, Quarterly Journal of Mathematics, vol. 36 (1904–5), pp. 373–384.

[9] H. R. Brahana, *Operators of order p^m in the group of isomorphisms of the abelian group of order p^n and type* 1, 1, 1 $\cdots$, Proceedings of the National Academy of Sciences, vol. 18 (1932), pp. 722–724.

easy to apply and allows a preliminary classification of subgroups which is probably the one natural to the problem since it leads to the application of methods which have been used in other mathematical investigations. The failure to use many of the powerful tools of mathematics has been one of the most striking features of the development of the theory of finite groups up to the present time.

By means of the theorem above the operators of I_p may be arranged in order of complexity. Let the partitions of n be arranged so that in each partition $n_i \geqq n_{i+1}$, and so that one partition precedes another if the first n_i in which they differ is smaller than the corresponding n_i in the other. An immediate justification for calling this an arrangement in the order of complexity is the fact that the group $\{H, U\}$ where U determines the partition $n = n_1 + n_2 + \cdots + n_s$ is of class n_1. Now every subgroup and every quotient group of a group of class k is of class not greater than k. Hence if we wish to investigate groups of class k we may disregard all subgroups of I_p which contain operators for which $n_1 > k$. Moreover, if G contains only operators of order p, as in the problem before us, and since if $n_1 \geqq p$ the corresponding operator U is of order at least p^2, we may suppose than $n_1 < p$.[10] For this reason the results obtained by Dickson[11] on the subgroups of I_p of low index do not apply in general, for p must be large with respect to n if a subgroup of index p or p^2 is to contain no operator of class greater than or equal to p.

Now since simple isomorphism of G and G' implies the simple isomorphism of all the quotient groups and in particular the identity of the two series of successive central quotient groups, and since the class of the central quotient group of G is one lower than the class of G, it follows that the non-isomorphism of G and G' may be determinable by the consideration of groups of very low class. The last central quotient group in each series is abelian and simple isomorphism of two such follows from the equality of their orders. The next to the last central quotient group is metabelian. Classification of metabelian groups is therefore the next indispensable step in an attack on the fundamental problem. For these groups the subgroups of I_p corresponding to G/H contain only operators for which $n_1 = 2$.

The subgroups of I_p which contain only operators for which $n_1 = 2$ are abelian. Two such subgroups are simply isomorphic if they have the same order. The operators are ordered above. We may say that U is *of type* k if

$$n_1 = n_2 = \cdots = n_k = 2$$

and $n_{k+1} = 1$ whenever $k < n/2$. Two subgroups of I_p may be conjugate only if they contain operators of the same types. For the discussion and classification of the abelian subgroups of I_p which contain only operators for which $n_1 = 2$ it will be convenient to consider the corresponding metabelian subgroups of the holomorph of H.

[10] This does not say that the class of G is less than p.

[11] Loc. cit., footnote 8.

The groups which we are considering, i.e., groups whose operators are all of order p, belong to the class which P. Hall has called[12] *regular* p-groups. Hall has determined many properties of regular p-groups and many invariants. He is interested mainly in what may be said of all p-groups. One looks for aid on the fundamental problem from investigations of that sort. The aid will come, however, only when the subject has progressed far enough to arouse interest in differences among groups and to offer information concerning complete sets of invariants. The large numbers of groups identical with respect to all the invariants so far determined suggest that that stage of development is not imminent.

2. The trilinear form determined by G. Every metabelian group G, composed of operators all except identity of order p, which contains H as a maximal invariant abelian subgroup and is generated by H and m permutable operators $U_1, U_2, \cdots, U_m$ from the group of isomorphisms of H determines a trilinear form with coefficients in a $GF(p)$

$$F(x, y, z) = \Sigma\, a_{hij}\, x_h y_i z_j .$$

Conversely, every such trilinear form determines a metabelian group having the above properties.[13] Two groups determined by two forms $F_1(x, y, z)$ and $F_2(x, y, z)$ are simply isomorphic if and only if it is possible to transform F_1 into F_2 by means of linear transformations with integer coefficients on the x's, the y's, and the z's separately. The problem of the classification of these groups is therefore precisely the problem of the determination of complete sets of invariants of the forms under rational linear transformations of the variables. Problems of this type, where the transformations are, however, more general and the coefficients are not in a modular field, have been studied for a long time by M. Lecat[14] and by L. H. Rice and Hitchcock.[15] More recently R. Oldenburger has announced[16] the determination of a complete set of invariants for trilinear forms. It is our purpose here to examine the relation between the groups and the forms and to interpret properties of one in terms of the other. In particular, a set of properties which is sufficient to determine a group G must, when re-interpreted, constitute a complete set of invariants for the form F. Definitive sets of properties for a comparatively large number of classes of groups

[12] *A contribution to the theory of groups of prime-power order*, Proceedings of the London Mathematical Society, (2), vol. 36 (1933), pp. 29–95.

[13] These statements are obvious re-interpretations of the results of my paper *Metabelian groups and pencils of bilinear forms*, to appear in the American Journal of Mathematics.

[14] *Coup d'oeil sur la théorie des déterminants supérieurs*, Brussels, 1927.

[15] Journal of Mathematics and Physics, Massachusetts Institute of Technology, vols. 4–8 (1925–9).

[16] *Canonical triples of bilinear forms*, Bulletin of the American Mathematical Society, vol. 40 (1934), p. 226.

have already been determined[17] so that we have at hand definitive information about many types of form.

It is necessary first to specify more completely the relation between G and F. Let G be generated by $H = \{s_1, s_2, \cdots, s_n\}$ and $U = \{U_1, U_2, \cdots, U_m\}$, where H and U are abelian and of type 1, 1, $\cdots$ and no operator of U is permutable with every operator of H. The fact that G is metabelian requires every operator of U to correspond to a partition of n whose greatest term is 2. Let the central C of G, composed of operators of H which are invariant under U, be of order p^c, and let $k = n - c$. We may suppose that generators of H are chosen so that the last c of them, $s_{k+1}, s_{k+2}, \cdots, s_n$, are in the central. Then $s_1, s_2, \cdots, s_k$ will be non-invariant in G. U can contain no operator of type greater than k, i.e., no operator of U can correspond to a partition of n in which there are more than k 2's. Let the commutator subgroup K of G be of order p^l. K is in C and the generators of C can be chosen so that the first l of them, $s_{k+1}, s_{k+2}, \cdots, s_{k+l}$, are in K and hence generate K.

Now the trilinear form $F(x, y, z)$ which has for a_{hij} the exponent of s_{k+i} in the commutator of s_h and U_j is completely determined by the choice of generators of K, H, and U. A different choice of generators would determine a different form F'. If $U_1', U_2', \cdots, U_m'$ were selected from U so that

$$U_i' = U_1^{\alpha_{i1}} U_2^{\alpha_{i2}} \cdots U_m^{\alpha_{im}},$$

F' would be obtained from F by applying the transformation

$$z_i' = \alpha_{i1} z_1 + \alpha_{i2} z_2 + \cdots + \alpha_{im} z_m .$$

Likewise if generators of K were selected differently, F' would be obtained from F by making the proper transformation on the variables y, and if generators of H were selected differently, F' would be obtained by making the proper transformations on the variables x.

3. **Invariants of** F. Some of the invariants of F under rational transformation on the variables are immediately obvious. It will be convenient to consider F as a three-way matrix, the matrix of coefficients. The terms of F for which h is constant constitute a two-way matrix of m rows and l columns; these matrices are k in number. We shall call these h-matrices. We may define i-matrices and j-matrices in an analogous manner. Now it is obvious that we can select a set of operators to generate G which satisfy all of the conditions imposed above except conditions of independence and that such a set of generators will determine a form F' with different numbers of h-, i-, and j-matrices, thus giving different numbers k', l', m' in place of k, l, m. However, the number l which is defined by the order of the commutator subgroup of G is an invariant of G and hence must be an invariant of F. This number is obviously the smallest number of i-matrices in terms of which the rest are linearly ex-

[17] Cf. the reference in footnote 13 and the references given in that paper.

pressible. Similarly, k and m are the smallest numbers of h- and j-matrices in terms of which the rest may be linearly expressed. Thus the smallest numbers of x's, y's, and z's in terms of which the form may be expressed are invariants of the form. These numbers are the exponents of p in the index of C in H, the order of K, and the order of U respectively.

For some forms these three invariants constitute a complete set. An obvious generalization of Theorem (1.1) of the paper *Metabelian groups and pencils of bilinear forms* gives

(3.1) *Two trilinear forms with the same invariants k, l, and m, where $km = l$, are conjugate.*[18]

Another invariant of G under the transformations which leave H invariant is the type of the most complex operator of U, i.e., the number of 2's in the partition of n determined by that operator. If U_1 is of type r, the commutator subgroup arising from transformation of H by U_1 is of order p^r. This requires that the rank of the j-matrix with $j = 1$ be r. If U contains operators of type r', then there exist linear combinations of j-matrices which are of rank r'. When $km - 1 = l$, the invariants described above are sufficient to determine the conjugate set to which the form belongs. Theorem (1.2) of the paper cited above is generalized and translated into

(3.2) *Two trilinear forms with the same invariants k, l, and m, where $km - 1 = l$, belong to the same conjugate set if both contain or do not contain j-matrices of rank $k - 1$. There exist forms of each of the two types.*

It is not necessary to follow the subject very far to see that the invariants described above are not sufficient in general to determine the conjugate set to which the form belongs. When $k = 2$ and $m = 3$, there exist two groups which require the determination of another invariant in order to distinguish between them. For these values of k and m we have $2 \leqq l \leqq 6$.[19] There are 17 types of form. Of these, 15 are determined by the invariants described above. For the other two we have $k = 2$, $l = 3$, $m = 3$, and each contains the same number of j-matrices of rank 1. They are distinguished by the fact that one contains i-matrices of rank 1, whereas the other contains only i-matrices of rank 2.

There is one other interesting thing obtained by translating the table referred to in footnote 19 into terms of the trilinear forms. There is a group determined by the facts that $k = 2$, $l = 3$, $m = 3$, and all of its j-matrices are of rank 2. The last requirement is equivalent to the condition that the determinant of

$$(a_{1ij}x_1 + a_{2ij}x_2),$$

which is a homogeneous cubic in x_1 and x_2, be irreducible in the $GF(p)$.

[18] We understand the term "conjugate" to mean "conjugate under rational linear transformations on the x's, the y's, and the z's".

[19] Cf. p. 510 of my paper *On the metabelian groups which contain a given group H as a maximal invariant abelian subgroup*, American Journal of Mathematics, vol. 56 (1934), pp. 490–510.

4. **The groups determined by a given form.** In setting up the trilinear form determined by a given group special rôles are assigned to the variables x, y, and z. The x's correspond to non-invariant operators in H, the y's correspond to operators in the commutator subgroup, and the z's correspond to operators in the group of isomorphisms of H. There is nothing special, however, in the ways in which the different sets of variables are treated. Looking at the form and the transformations to which it is subjected, there is no apparent reason why y's should correspond to commutators rather than x's, or z's. The form actually describes a set of defining relations which are satisfied by generators of the group.[20] It is obvious that for any given form and any interpretation of the x's, y's, and z's, a set of defining relations may be given and a group thereby determined. There are six possible interpretations of the x's, y's, and z's in terms of non-invariant operators of H, operators of K, and operators of U. There are therefore six possible groups determined by a given form. When k, l, and m are distinct numbers at least three of the six groups are distinct, for the commutator subgroup is a characteristic subgroup and two groups with commutator subgroups of orders p^k and p^l are not simply isomorphic when $k \neq l$. Consequently if a set of invariants sufficient to determine the conjugate set to which a form belongs is given, this set of invariants can in general be translated in more than one way into a set of quantities which will determine a group.

We shall show first that not more than three of the six possible groups are distinct. It is quite easy to see that the rôles of the x's and the z's are interchangeable. The group $G = \{H, U\}$ contains the maximal invariant abelian subgroup $H = \{C, s_1, s_2, \cdots, s_k\}$ and the maximal invariant abelian subgroup $H' = \{C, U_1, U_2, \cdots, U_m\}$. The operators $s_1, s_2, \cdots, s_k$ correspond to k operators of order p in the group of isomorphisms of H'. If a new trilinear form F' were constructed corresponding to the second maximal invariant abelian subgroup, we should then have F except that the x's would be called z's, and vice versa. Thus if from a given form we determine two groups by interchanging the rôles of the x's and the z's, the two groups are simply isomorphic.

Whenever k, l, and m are distinct, there are three different groups determined by a given form. The last paragraph shows that in considering the groups we may suppose that k is never greater than m. Moreover, if k, l, and m are not equal and we know enough invariants of the form to determine the conjugate set to which it belongs, we have enough information to determine at least two distinct groups, and enough to determine three, if k, l, and m are distinct. Now if we have enough information to determine one group, this information translated into invariants of the form is enough to determine the conjugate set to which the form belongs and hence translates back to determine one or two new

[20] The defining relations in question do not strictly determine the group because neither n nor c enters directly into the considerations; the difference $n - c$ does enter. Two groups which determine the same form with the same interpretation of x, y, and z are such that one is the direct product of the other by an abelian group of type $1, 1, \cdots$.

groups according as k, l, and m have two or three distinct values. *The consideration of the form has therefore cut down to about one-third the amount of work required to classify groups of this type.*

5. **Some illustrative examples.** Let us use the symbol $[k, l, m]$ to designate a trilinear form in k x's, l y's, and m z's. A complete set of invariants for $[k, l, m]$ will be a complete set for $[k, m, l]$, $[l, k, m]$, etc. We shall take a brief look at the groups determined by $[2, 6, 3]$. From Theorem (3.1) it follows that the conjugate set to which each of the forms belongs is determined by the three numbers used in the symbol for it. $[2, 6, 3]$ determines a group G generated by operators $s_1, \cdots, s_n, U_1, U_2, U_3$ which are independent except for relations which are consequences of

$$U_1^{-1}s_1U_1 = s_1s_3, \qquad U_2^{-1}s_1U_2 = s_1s_5, \qquad U_3^{-1}s_1U_3 = s_1s_7,$$
$$U_1^{-1}s_2U_1 = s_2s_4, \qquad U_2^{-1}s_2U_2 = s_2s_6, \qquad U_3^{-1}s_2U_3 = s_2s_8.$$

Every j-matrix is of rank 2 corresponding to the fact that every operator of U is of type 2. Every h-matrix is of rank 3, since every operator of $\{s_1, s_2\}$ gives rise to a commutator subgroup of order p^3 when transformed by U. The i-matrices are all of ranks 1 and 2.

Corresponding to the form $[3, 6, 2]$ we have the group G' generated by operators which satisfy the relations

$$U_1^{-1}s_1U_1 = s_1s_4, \qquad U_2^{-1}s_1U_2 = s_1s_7,$$
$$U_1^{-1}s_2U_1 = s_2s_5, \qquad U_2^{-1}s_2U_2 = s_2s_8,$$
$$U_1^{-1}s_3U_1 = s_3s_6, \qquad U_2^{-1}s_3U_2 = s_3s_9.$$

In this case every j-matrix is of rank 3, corresponding to the fact that every operator of U is of type 3. Every h-matrix is of rank 2, corresponding to the fact that every operator of $\{s_1, s_2, s_3\}$ gives rise to a commutator subgroup of order p^2 when transformed by U. The i-matrices are the same as those for $[2, 6, 3]$ except that rows have become columns and columns have become rows. As was pointed out in the last section, these two groups are the same.

Let us now consider the form $[2, 3, 6]$. Corresponding to this form there exists a group G'' generated by operators which satisfy the relations

$$U_1^{-1}s_1U_1 = s_1s_3, \qquad U_2^{-1}s_1U_2 = s_1s_4, \qquad U_3^{-1}s_1U_3 = s_1s_5,$$
$$U_4^{-1}s_2U_4 = s_2s_3, \qquad U_5^{-1}s_2U_5 = s_2s_4, \qquad U_6^{-1}s_2U_6 = s_2s_5.$$

The j-matrices corresponding to $U_1, \cdots, U_6$ are all of rank 1. This reflects the fact that each of the operators $U_1, \cdots, U_6$ is of type 1. The group U, however, contains operators of type 2, and of course there are linear combinations of the j-matrices which are of rank 2. The h-matrices are all of rank 3, corresponding to the fact that every operator of $\{s_1, s_2\}$ gives rise to a commutator

subgroup of order p^3 when transformed by U. The i-matrices are all of rank 2. This group is one that has not appeared in any of the lists that have been made.[21]

In order to illustrate the advantages accruing from the considerations of the last section we shall examine this group by the methods used to obtain the lists referred to above. Leaving out the information coming from a consideration of the trilinear form it is not obvious that there is but one group $G = \{H, U\}$ of order p^{n+6} with commutator subgroup of order p^3 and central of order p^{n-2}. However, since K is of order p^3, at least three U's must be permutable with s_2 and hence are of type 1. And also there exist three U's which are permutable with s_1 and are therefore of type 1. Since H is maximal abelian none of the three U's which are permutable with s_1 can be permutable with s_2 and hence these six U's generate U. If we denote the three U's which are permutable with s_2 by U_1, U_2, U_3 respectively, and the three which are permutable with s_1 by U_4, U_5, U_6, it is obvious that operators U_4', U_5', U_6' can be selected in $\{U_4, U_5, U_6\}$ such that the commutator of U_{3+i}' and s_2 is the same as that of U_i and s_1. The number of subgroups of order p composed of operators of type 1 in U can be found quite easily. There are $1 + p + p^2$ in each of $\{U_1, U_2, U_3\}$ and $\{U_4, U_5, U_6\}$ and there are also the groups generated by $U_1^\alpha U_2^\beta U_3^\gamma U_4^\alpha U_5^\beta U_6^\gamma$, where α, β, γ gives the same group as $k\alpha$, $k\beta$, $k\gamma$, $k \neq 0$. There are $1 + p + p^2$ of these groups and they are distinct from those described above. There are therefore $3(1 + p + p^2)$ subgroups of U composed of operators of type 1.

We now consider the group corresponding to [3, 2, 6]. It is generated by operators which satisfy the relations

$$U_1^{-1}s_1U_1 = s_1s_4, \qquad U_2^{-1}s_2U_2 = s_2s_4, \qquad U_3^{-1}s_3U_3 = s_3s_4,$$

$$U_4^{-1}s_1U_4 = s_1s_5, \qquad U_5^{-1}s_2U_5 = s_2s_5, \qquad U_6^{-1}s_3U_6 = s_3s_5.$$

The j-matrices are all of ranks 1 and 2, being the same as those in the preceding case. The h-matrices are of rank 2 and the i-matrices are of rank 3. This group G''' is of order p^{n+6} with K of order p^2 and C of order p^{n-3}.

6. **Relations between two groups determined by the same form.** The situation described in §4 explains a question that arose in an earlier paper but was not investigated at that time.[22] In discussing abelian subgroups of I_p, it was noted that any such subgroup containing operators of type 2 and none more complicated was composed of operators of

$$\begin{bmatrix} 1 & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & 1 & a_{23} & \cdots & a_{2n} \\ 0 & 0 & 1 & \cdots & a_{3n} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

[21] Complete lists have been made only for $k = 2$ and $m \leqq 4$.

[22] Loc. cit., footnote 19, p. 493.

with every $a_{ij} = 0$ for $i > 2$, or else of operators with every $a_{ij} = 0$ for $j < n - 2$. The two possible U's so obtained would give different G's because in general their commutator subgroups would have different orders. It was recognized that there were two possible ways of going forward and one was selected. It is now clear that if the other had been selected the same trilinear form would have been reached except that x's and y's would have been interchanged, for that interchange does not affect ranks of j-matrices which determine the types of operators in U. Groups G'' and G''' of §5 illustrate this relation.

Another question which arose is illuminated here. The classifications were carried out by following closely the question of types of operators in U. This is a question of ranks of j-matrices. Before that aspect was fastened upon and pursued relentlessly an attempt was made to classify the groups by considering commutator subgroups of $\{C, s_1, U\}$ and $\{C, s_2, U\}$. The considerations became intricate and were abandoned. This was before it was thoroughly appreciated that the considerations were necessarily intricate. It is now clear that the method was perfectly feasible and is the method finally followed except that ranks of i-matrices were being considered instead of ranks of j-matrices.

7. Further relations between G and F. A j-matrix has been seen to describe a set of generating relations of a subgroup $\{H, U_j\}$ of G. Likewise an h-matrix describes a set of generating relations of a subgroup $\{C, s_h, U\}$. An i-matrix, however, does not describe a subgroup of G. If the particular i-matrix is the one corresponding to $i = i_1$, it may be obtained from F by setting $a_{hij} = 0$ for $i = 1, 2, \cdots, i_1 - 1, i_1 + 1, \cdots, l$. It describes a set of generating relations of the quotient group of G with respect to the invariant subgroup $\{s_{k+1}, s_{k+2}, \cdots, s_{k+i_1-1}, s_{k+i_1+1}, \cdots, s_{k+l}\}$. If the given i-matrix is of rank 1, the quotient group defined by it has an abelian subgroup of index p, for generators of the quotient group G' correspond to generators of G and pairs of generators of G whose commutators have been set equal to identity determine permutable generators of G'. The fact that the i-matrix is of rank 1 says that all the columns are linearly dependent on one column or all the rows on one row. By a proper choice of generators $s_1', s_2', \cdots, s_k', U_1', \cdots, U_m'$ the i-matrix is reduced to one with a single non-zero element. There is therefore just one pair of non-permutable generators of G'.

It is also clear that any set of h-matrices or any set of j-matrices determines a subgroup of G and any set of i-matrices determines a quotient group of G. Normal forms for generating relations of G will of course be concerned with the kinds of subgroup contained in G and the kinds of groups to which G is multiply isomorphic. Suppose a set of m' operators $U_1', U_2', \cdots, U_{m'}'$ is selected from U and consider the group G' generated by H and U'. G' determines a form F' which is composed of m' j-matrices of F. Though this form has k x's and l y's, it is not certain that k' and l' are the same as k and l. The h-matrices of F are linearly independent but the h-matrices of F' are not necessarily so.

Questions of that kind had to be considered in obtaining a set of normal forms for the groups.

By way of illustration let us consider the forms [2, 6, 4]. There are seven conjugate sets of such forms.[23] Of these, three have j-matrices all of rank 2, two have each one j-matrix of rank 1,[24] one has two j-matrices of rank one, and one has $1 + p$ j-matrices of rank 1. The two which contain one j-matrix of rank 1 are distinguished by the fact that one contains a form [2, 5, 3] with j-matrices all of rank 2 and the other contains no such form. The three with no j-matrices of rank 1 are distinguished by the facts that one contains a form [2, 2, 2]; the second contains a form [2, 4, 3] and no [2, 2, 2]; the third does not contain either.

There are $p + 2$ forms [2, 4, 4], each of which has all its j-matrices of rank 2. We shall not give a complete classification of these, but shall point out one invariant that has not appeared before. Each of $(p + 1)/2$ of them contains no form [2, 2, 2]. These are distinguished by the cross-ratio of the zeros of the determinant

$$| a_{1ij}x_1 + a_{2ij}x_2 | \qquad (i, j = 1, \cdots, 4),$$

which is irreducible.

8. **Concluding remarks.** At this point the student of finite groups is pulled in several different directions. Having presented the problem to the students of forms, he may await its solution with confidence. However, knowing something of the properties that are useful in characterizing a group he may wish to take a hand in the selection of the invariants that shall be used to characterize a conjugate set of forms. Also, all those forms of type $[k, l, l]$ are intriguing. For these the determinant of

$$(a_{1ij}x_1 + a_{2ij}x_2 + \cdots + a_{kij}x_k)$$

does not vanish identically, and the problem of the classification of the groups is the same as that of the classification of $(k - 2)$-dimensional surfaces of order l in a finite projective space of $k - 1$ dimensions. The forms [3, 3, 3] require the study of plane cubic curves. It will be interesting to see how the significant properties of plane curves appear in the theory of groups. The forms $[2, m, m]$ have been studied to some extent[25] but promise to repay further study. Is it impossible to distinguish one from another the groups at the end of §7 by any consideration of their subgroups or quotient groups?

There is also the question of abelian subgroups of I_p that are not restricted to operators for which $n_1 = 2$. These lead to subgroups of the holomorph of H

[23] They correspond to groups numbered 4, $\cdots$, 10 in the paper cited in footnote 13.

[24] Here the matrix (α_{ij}) is counted the same as $(d\alpha_{ij})$.

[25] H. R. Brahana, *Metabelian groups of order p^{n+m} with commutator subgroups of order p^m* Transactions of the American Mathematical Society, vol. 36 (1934), pp. 776–792.

which are not metabelian, but which may lend themselves to a simpler treatment than the general group of class k.[26]

The student who keeps the fundamental problem before him will be anxious to investigate subgroups of class 3 of the holomorph of H. He will try to discover or devise machinery as suitable as the trilinear form proved to be for the metabelian subgroups. Considering the machinery used thus far in its geometric form, one wonders what is left to cope with the increased complexity of groups of class 3. But if the student wishes to guard against an attack from the rear he must first investigate the metabelian groups which are not subgroups of the holomorph of H. This we feel called upon to do.

URBANA, ILLINOIS.

[26] This problem is attacked in Miss Henrietta Terry's paper *Abelian subgroups of order* p^m *of the I-groups of the abelian groups of order* p^n *and type* 1, 1, 1, $\cdots$, this Journal, vol. 1 (1935), pp. 27–34.

NATURAL ISOPERIMETRIC CONDITIONS IN THE CALCULUS OF VARIATIONS

By G. D. Birkhoff and M. R. Hestenes

Introduction

In the ordinary problems of the Calculus of Variations, an extremal is defined in general as an arc along which the first variation of the prescribed integral J vanishes. Such an arc is completely characterized by the fact that it satisfies the attached Euler differential equations and transversality conditions.

Among such extremal arcs there are certain ones which furnish a minimum value of J under the prescribed boundary conditions. It has always been a matter of primary theoretical interest in the Calculus of Variations to determine when such a minimum is realized. In the case of a minimum the minimizing arc can be found by use of a minimizing principle.

For extremal arcs not furnishing a minimum the attached quadratic accessory minimum problem no longer is analogous to a positive definite quadratic form. It has long been recognized that the number of negative terms, or 'type number', in this normalized form is equal to the number of negative characteristic values in the associated linear self-adjoint boundary value problem. This fact has been especially apparent ever since the formulation of the theory of homogeneous linear integral equations with real symmetric kernels. It is implicitly involved in the work of Birkhoff[1] without regard to the boundary value problem, in particular for the case of type number zero (minimum) and type number one (minimax). Morse[2] has subsequently developed the notion of type number systematically, also using the method of broken extremals.

The question arises as to whether or not all extremals whatsoever may not be obtained as a solution of a properly formulated minimum problem. In the present paper we show that this is indeed the case. More definitely, we show that by adding a suitable set of 'natural isoperimetric conditions', *which are automatically satisfied by every possible extremal fulfilling the given conditions*, there is obtained a related isoperimetric problem for which the extremal arc in question is a minimizing arc.[3]

So far as we know, the only case in which such natural isoperimetric condi-

Received March 25, 1935.

[1] See his paper *Dynamical systems with two degrees of freedom*, Trans. Amer. Math. Soc., vol. 18 (1917), in particular, pp. 239–257.

[2] For references to the works of Morse, see his book *The Calculus of Variations in the Large*, Colloquium Publications, American Mathematical Society, vol. 18 (1934). Unless otherwise expressly stated, all references to Morse in the following pages are to his book.

[3] See our paper in the Proceedings of the National Academy of Sciences, February, 1935, with the same title, in which this general principle is first formulated.

tions have been used similarly is in a paper by Poincaré[4] in which he proves the existence of a closed geodesic on any closed convex analytic surface. It is obvious that such a geodesic is not of minimum length. However, by adjoining the condition that admissible curves divide the surface in question into two parts of equal integral curvature, he proved that there exists a shortest curve of this type which is the desired closed geodesic. The 'natural isoperimetric condition' here entering is automatically fulfilled along all closed geodesics and may be reduced to one of the kind which we consider. It may be noted, incidentally, that the type number of such a minimizing geodesic is *one*, corresponding to the fact that only a single condition has been imposed.

In the present paper we define first such natural isoperimetric conditions as seem to be of the most useful type. Conditions of more general type are described later. Furthermore, we have not endeavored to treat the difficult question of the existence of natural isoperimetric conditions valid in the large.

With our approach the type number appears naturally as the minimum number of such natural isoperimetric conditions which suffice to reduce the given problem to a minimizing problem. It is then proved subsequently that the definition so obtained agrees with that referred to above. At least in the case of one dimension this attack renders unnecessary all reference to the accessory boundary value problem. It has the further advantage of remaining wholly in the classical domain of the Calculus of Variations and of avoiding the use of the method of broken extremals.

It might be expected that the isoperimetric problem arising on the adjunction of such conditions would require the usual isoperimetric theory for its solution. It is shown here that this is not the case, so that the theory can be made to depend upon the ordinary theory of the minimum whether in parametric or non-parametric form.

In the proof of the minimizing property here given an interesting lemma in the ordinary Calculus of Variations is established; namely, that if the ordinary conditions for a minimum in the parametric case with fixed end points is satisfied along an extremal arc PQ, then, *even if all arcs PQ in the neighborhood of any segment of the given extremal, no matter how far extended, be admitted,* the arc PQ will still furnish the absolute minimum in the extended neighborhood. A further interesting result is that in the variable end point case a non-degenerate extremal arc E_{12} can be reduced to a minimizing arc by a suitably chosen set of natural isoperimetric conditions even if its end points are conjugate.

Naturally enough our attack upon the general problem is made by considering first the allied accessory problem involving certain accessory natural isoperimetric conditions. In this manner it is possible to obtain the oscillation and comparison theorems of Morse in more complete form, as well as a new oscillation theorem in §6. It may be remarked that all of these results correspond to rather evident facts about ordinary quadratic forms and from this

[4] Trans. Amer. Math. Soc., vol. 6 (1905), pp. 237-274. In this connection, however, see the latter part of §13.

point of view can be predicted heuristically. As an instance of this parallelism we cite the following. If we vary the coefficient of a homogeneous quadratic form from one of positive definite type into another not of positive definite type, the associated linear system of equations for the characteristic values will vary likewise and k negative roots will be introduced if the final type number is k. A simple illustration of such an application in the one-dimensional case is obtained by allowing the interval to expand continuously, in which case a characteristic value appears with each successive conjugate point.

As is demonstrated in the present paper, the theory of natural isoperimetric conditions extends to problems in higher dimensions. A typical two-dimensional case is considered; but in order to treat it we have used the accessory boundary value problem; and we have limited ourselves principally to this accessory problem. It is to be hoped that ultimately the use of the accessory boundary value problem can be avoided, so that sufficient as well as necessary conditions for a strong minimum can be independently obtained, at least in simple cases. If this were accomplished, it would be possible to apply the Calculus of Variations directly to the discussion of the solutions of linear self-adjoint differential systems under self-adjoint boundary conditions.

For the case of higher dimensions, it is obvious that the use of natural isoperimetric conditions is more satisfactory than that of broken extremals which have no simple analogue in the case of multiple integrals.

I

The accessory problem in the fixed end point case[5]

In §§1–6 we shall study the integral

$$J(\eta) = \int_{x_1}^{x_2} 2\omega(x, \eta, \eta')\, dx$$

in $(x, \eta_1, \cdots, \eta_n)$-space subject to the end conditions

$$\eta_i(x_1) = 0, \qquad \eta_i(x_2) = 0 \qquad (i = 1, \cdots, n),$$

where[6]

$$2\omega = P_{ik}(x)\eta_i\eta_k + 2Q_{ik}(x)\eta_i\eta_k' + R_{ik}(x)\eta_i'\eta_k' \qquad (i, k = 1, \cdots, n).$$

This problem arises in the study of the second variation for the fixed end point case and is usually referred to as the accessory problem. A complete discussion of this problem is essential for the study of the more complicated problems discussed in this paper. The principal result obtained in these sections is that of showing that if the strengthened condition of Legendre holds, then the mini-

[5] The various parts of this paper need not be read consecutively. If one is interested only in the fixed end point case, it is sufficient to read parts I and III. These parts give the basic principles underlying our work.

[6] Here and elsewhere a repeated index denotes summation with respect to that index.

mum number of natural isoperimetric conditions needed to reduce the arc $(\eta) \equiv (0)$ to a minimizing arc is equal to the sum of the orders of the conjugate points of x_1 between x_1 and x_2.

1. **Preliminary definitions and theorems.** A function $f(x_1, \cdots, x_n)$ will be said to be of *class* C^n if it is continuous and possesses continuous derivatives of the first n orders. We shall assume that the functions $P_{ik} = P_{ki}$ are continuous and that the functions Q_{ik}, $R_{ik} = R_{ki}$ are of class C' on the interval x_1x_2. We admit functions $\eta_i(x)$ which are continuous and are composed of a finite number of segments of class C'. Such functions will be said to be of *class* D'. A set of functions $\eta_i(x)$ of class D' which vanish at x_1 and x_2 will be called a *set of admissible variations* for the fixed end point problem here considered.

The Euler equations of the integral $J(\eta)$ are the equations

$$L_i(\eta) = \omega_{\eta_i} - (d/dx)\omega_{\eta'_i} = 0 \qquad (i = 1, \cdots, n), \tag{1.1}$$

and are commonly called the *accessory* or *Jacobi equations*. A solution $\eta_i(x)$ of these equations of class C'' will be called an *accessory extremal*. From the theory of linear differential equations of the second order it follows that if the determinant $| R_{ik}(x) |$ is different from zero on x_1x_2, then an extremal is uniquely determined by the values of η_i, η'_i or of η_i, $\zeta_i = \omega_{\eta'_i}$ at a single point x_0.

A value $x_3 \neq x_1$ will be said to be *conjugate* to x_1 if there exists an accessory extremal η_i having $\eta_i(x_1) = \eta_i(x_3) = 0$ and $(\eta) \not\equiv (0)$ on x_1x_3. By the order of x_3 as a conjugate point of x_1 will be meant the maximum number of linearly independent accessory extremals having $\eta_i(x_1) = \eta_i(x_3) = 0$.

It is well-known that if the determinant $| R_{ik} |$ is different from zero, then there exists a set of $2n$ linearly independent accessory extremals $\eta_{is}(x)$ $(s = 1, \cdots, 2n)$ and that every accessory extremal is expressible linearly with constant coefficients in terms of these $2n$ extremals. Moreover, the points x_3 conjugate to x_1 are the zeros $x_3 \neq x_1$ of the determinant

$$\begin{vmatrix} \eta_{is}(x) \\ \eta_{is}(x_1) \end{vmatrix} \qquad (s = 1, \cdots, 2n).$$

It follows that if the point x_2 is not conjugate to x_1, then every pair of points (x_1, η_{i1}), (x_2, η_{i2}) can be joined by a unique extremal arc. Furthermore, it is clear that there exists a set of n linearly independent accessory extremals $\eta_{ik}(x)$ $(k = 1, \cdots, n)$ satisfying the conditions $\eta_{ik}(x_1) = 0$ and that every extremal η_i having $\eta_i(x_1) = 0$ is expressible linearly with constant coefficients in terms of these n extremals. The conjugate points x_3 of x_1 are then also determined by the zeros of the determinant $D(x) = | \eta_{ik}(x) |$.

It should be noted also that if the determinant $| R_{ik} |$ is different from zero, and $\eta_i(x)$ is a solution of equations (1.1) of class D' such that the functions $\zeta_i = \omega_{\eta'_i}$ are continuous on x_1x_2, then $\eta_i(x)$ is necessarily an accessory extremal.

2. Natural isoperimetric conditions. Let $\xi_i(x)$ be an admissible variation of class C'' and let

$$\Omega(\xi, \eta) = \xi_i \omega_{\eta_i} + \xi'_i \omega_{\eta'_i}. \tag{2.1}$$

The condition

$$J(\xi, \eta) = \int_{x_1}^{x_2} \Omega(\xi, \eta)\, dx = 0$$

will be called a *natural isoperimetric condition* for the functional $J(\eta)$. *It is satisfied by every accessory extremal* η_i, since in this case we have by integration by parts with the help of equation (1.1)

$$J(\xi, \eta) = \int_{x_1}^{x_2} (d/dx)(\xi_i \omega_{\eta'_i})\, dx = \xi_i \omega_{\eta'_i} \Big|_{x_1}^{x_2} = 0.$$

Consider now a set of m natural isoperimetric conditions

$$J(\xi_\alpha, \eta) = 0 \qquad (\alpha = 1, \cdots, m). \tag{2.2}$$

The Euler equations of $J(\eta)$ subject to these conditions are the equations

$$L_i(\eta + \mu_\alpha \xi_\alpha) = 0, \qquad \mu'_\alpha = 0 \qquad (\alpha = 1, \cdots, m). \tag{2.3}$$

A solution η_i, μ_α of class C'' of these equations will be called an *isoperimetric extremal.*

The set of conditions (2.2) will be called a *proper set* of natural isoperimetric conditions if the determinant

$$| J(\xi_\alpha, \xi_\beta) |$$

is different from zero. Concerning such sets we have the following interesting theorem:

THEOREM 2.1. *If the set* (2.2) *forms a proper set of natural isoperimetric conditions, then the multipliers* μ_α *belonging to an isoperimetric extremal* η_i, μ_α *satisfying the conditions* (2.2) *are all zero.*

For let η_i, μ_α be an isoperimetric extremal satisfying the conditions (2.2). By the use of equations (2.3) and integration by parts it is found that

$$0 = J(\xi_\alpha, \eta) = \int_{x_1}^{x_2} \{\xi_{i\alpha} L_i(\eta) + (d/dx)(\xi_{i\alpha}\omega_{\eta'_i})\}\, dx$$

$$= -\int_{x_1}^{x_2} \xi_{i\alpha} L_i(\mu_\beta \xi_\beta)\, dx = -J(\xi_\alpha, \xi_\beta)\mu_\beta.$$

It follows that $\mu_\beta = 0$ and the theorem is established.

We have further

THEOREM 2.2. *If* $\eta_{i0}(x)$ *is an admissible variation such that the equation*

$$J(\eta_0, \eta) = 0 \tag{2.4}$$

is true for every set of admissible variations η_i of class C'' satisfying the conditions (2.2), *then there exists a set of constants μ_β, c_i such that the equations*

$$\omega_{\eta_i'} = \int_{x_1}^{x} \omega_{\eta_i}\, dx + c_i \tag{2.5}$$

hold on x_1x_2 with

$$\eta_i = \eta_{i0} + \mu_\beta \xi_{i\beta}.$$

For let $\eta_{i\nu}(x)$ ($\nu = 1, \cdots, m + 1$) be a set of $m + 1$ admissible variations of class C''. If the hypotheses of the theorem are satisfied, then it is clear that the two matrices

$$\left\| \begin{matrix} J(\eta_0, \eta_\nu) \\ J(\xi_\alpha, \eta_\nu) \end{matrix} \right\|, \qquad \| J(\xi_\alpha, \eta_\nu) \|$$

must have the same rank for every set of $m + 1$ variations $\eta_{i\nu}$. It follows that there exists a set of constants μ_α such that

$$J(\eta_0, \eta) + \mu_\alpha J(\xi_\alpha, \eta) = J(\eta_0 + \mu_\alpha \xi_\alpha, \eta) = 0$$

for every admissible arc η_i of class C''. In fact this equation must hold for every admissible arc since every such arc is a limit curve of some sequence of admissible arcs of class C''. The theorem now follows from the fundamental lemma in the Calculus of Variations.

COROLLARY 1. *Suppose that the determinant $|R_{ik}|$ is different from zero on x_1x_2 and the set* (2.2) *forms a proper set of natural isoperimetric conditions for $J(\eta)$. Let $\eta_{i0}(x)$ be a set of admissible variations satisfying the conditions* (2.2). *If the equation* (2.4) *holds for every admissible arc η_i of class C'' satisfying the conditions* (2.2), *then η_{i0} is an accessory extremal.*

For, according to the remark at the end of the last section, the non-vanishing of the determinant $|R_{ik}|$ implies that the arc $\eta_{i0} + \mu_\alpha \xi_{i\alpha}$ is an extremal arc, since it satisfies the conditions (2.5). The corollary now follows from Theorem 2.1.

COROLLARY 2. *If the determinant $|R_{ik}|$ is different from zero on x_1x_2, then an admissible arc $\eta_{i0}(x)$ which satisfies all natural isoperimetric conditions is necessarily an accessory extremal.*

This result follows from Corollary 1 since in this case the condition (2.4) must hold for all admissible variations $\eta_i = \xi_i(x)$ of class C''.

3. **Minimal sets.** A set of natural isoperimetric conditions

$$J(\xi_\alpha, \eta) = 0 \qquad (\alpha = 1, \cdots, m) \tag{3.1}$$

will be called a *minimal set*, if the inequality $J(\eta) \geqq 0$ is true for every set of admissible variations η_i satisfying the conditions (3.1) and if no proper subset of these conditions has this property. One readily verifies that if the set

(3.1) is a minimal set, so also is the set $J(\bar{\xi}_\alpha, \eta) = 0$, where $\bar{\xi}_{i\alpha} = \xi_{i\beta}A_{\alpha\beta}$ and $|A_{\alpha\beta}| \neq 0$, the A's being constants.

THEOREM 3.1. *If the set* (3.1) *is a minimal set, then the inequality*

$$J(\xi_\alpha, \xi_\beta)a_\alpha a_\beta \leqq 0 \qquad (\alpha, \beta = 1, \cdots, m)$$

is true for every set of constants $(a) \neq (0)$.

For we may suppose that the set (3.1) has been chosen so that

$$J(\xi_\alpha, \xi_\beta) = 0 \qquad (\alpha \neq \beta). \tag{3.2}$$

Suppose now one of the numbers $J(\xi_\alpha, \xi_\alpha)$, say $J(\xi_m, \xi_m)$, were positive. Let $\eta_i(x)$ be an admissible variation satisfying the first $m - 1$ conditions (3.1), having $J(\xi_m, \eta) \neq 0$ and $J(\eta) < 0$, and choose a constant b such that

$$J(\xi_m, \eta) = bJ(\xi_m, \xi_m) = bJ(\xi_m).$$

The arc $\eta_i - b\xi_{im}$ would then satisfy the conditions (3.1) and have

$$\begin{aligned} J(\eta - b\xi_m) &= J(\eta) - 2bJ(\xi_m, \eta) + b^2J(\xi_m) \\ &= J(\eta) - b^2J(\xi_m) < 0, \end{aligned}$$

contrary to our assumption that the set (3.1) is a minimal set. This proves Theorem 3.1.

COROLLARY. *If the set* (3.1) *forms a proper minimal set, then the inequality*

$$J(\xi_\alpha, \xi_\beta)a_\alpha a_\beta < 0 \qquad (\alpha, \beta = 1, \cdots, m) \tag{3.3}$$

is true for every set of constants $(a) \neq (0)$. *Conversely, if the set* (3.1) *forms a maximal set of isoperimetric conditions such that the condition* (3.3) *holds, then the set* (3.1) *is a proper minimal set.*

The following theorem shows the importance of proper minimal sets.

THEOREM 3.2. *If the set* (3.1) *forms a minimal set, there exists a proper minimal set composed of the same number of natural isoperimetric conditions.*

For suppose the condition (3.2) holds and suppose $J(\xi_m, \xi_m) = 0$. Let ξ_i be an admissible arc satisfying the first $m - 1$ conditions (3.1) and having $J(\xi) < 0$. The set

$$J(\xi, \eta) = 0, \qquad J(\xi_\alpha, \eta) = 0 \qquad (\alpha = 1, \cdots, m - 1) \tag{3.4}$$

can now be shown to be a minimal set except for the fact that ξ_i may not be of class C''. For if there were an admissible arc η_i satisfying these conditions and having $J(\eta) < 0$, then by choosing a constant b such that

$$J(\xi_m, \eta) = bJ(\xi_m, \xi)$$

it would be found that the arc $\eta_i - b\xi_i$ would satisfy the conditions (3.1) and would have

$$J(\eta - b\xi) = J(\eta) - 2bJ(\xi, \eta) + b^2J(\xi) = J(\eta) + b^2J(\xi) < 0.$$

This is impossible. It remains to show that none of the conditions (3.4) can be dropped. Suppose the last one could be dropped and let η_i be an admissible arc having $J(\eta) < 0$ and satisfying the conditions (3.1) except for the $(m-1)$ st. By choosing a constant b so as to satisfy the equations

$$J(\xi, \eta) = bJ(\xi, \xi_m)$$

it would be found that the arc $\eta_i - b\xi_{im}$ would have $J < 0$ and would satisfy the equations (3.4) except for the last. The set (3.4) is accordingly a minimal set. By a repetition of this process we can eliminate all the numbers $J(\xi_\alpha, \xi_\alpha)$ which are zero and obtain thereby a proper minimal set as described in the theorem, since the functions $\xi_{i\alpha}$ as obtained which are not of class C'' can be replaced by functions of class C''.

In view of this result we may restrict ourselves to proper minimal sets when questions concerning minimal sets arise.

THEOREM 3.3. *The number of conditions in a minimal set is the same for every such set.*

For if the sets

$$J(\xi_\alpha, \eta) = 0 \qquad (\alpha = 1, \cdots, m_1),$$

$$J(\eta_\beta, \eta) = 0 \qquad (\beta = 1, \cdots, m_2)$$

form two proper minimal sets and m_2 were greater than m_1, then there would exist constants a_β not all zero such that

$$J(\xi_\alpha, \eta_\beta)a_\beta = 0.$$

The arc $\eta_i = \eta_{i\beta}a_\beta$ would then satisfy the first of these sets and have

$$J(\eta) = J(\eta_\beta, \eta_\gamma)a_\beta a_\gamma < 0 \qquad (\beta, \gamma = 1, \cdots, m_2)$$

by the corollary of Theorem 3.1. But this is impossible. Hence $m_2 \leqq m_1$, and the theorem is immediate.

The number of conditions in a minimal set will be called the *type number* of the integral J. If no minimal set exists for J, then we shall say its type number is infinite. In view of Theorem 3.3 it is clear that the type number of J is a well defined concept.

The following theorem gives a second method of evaluating the type number of J.

THEOREM 3.4. *Let $\eta_{i\beta}$ $(\beta = 1, \cdots, m)$ be a set of admissible variations satisfying the condition*

$$J(\eta_\alpha, \eta_\beta)a_\alpha a_\beta < 0 \qquad (\alpha, \beta = 1, \cdots, m) \tag{3.5}$$

for every set of constants $(a) \neq (0)$. The number of variations in a maximal set of variations of this type is equal to the type number of J.

This follows because if $\eta_{i\beta}$ forms a maximal set of variations satisfying the

condition (3.5), then we can replace these functions by admissible variations $\xi_{i\beta}$ of class C'' satisfying the same condition. The conditions

$$J(\xi_\beta, \eta) = 0 \qquad (\beta = 1, \cdots, m)$$

would then form a minimal set by the corollary to Theorem 3.1.

COROLLARY. *If the condition* (3.5) *holds for a set of admissible variations* $\eta_{i\beta}$ $(\beta = 1, \cdots, m)$, *the type number of* J *is at least* m.

The *condition of Legendre* will be said to hold on x_1x_2 if the inequality

$$R_{ik}(x)\pi_i\pi_k \geqq 0 \qquad (i, k = 1, \cdots, n) \tag{3.6}$$

is true for every set of constants $(\pi) \neq (0)$. If the condition (3.6) holds with the equality sign excluded, the *strengthened condition of Legendre* will be said to hold on x_1x_2.

THEOREM 3.5. *If the type number* m *of* J *is finite, the condition of Legendre holds on* x_1x_2. *Moreover, if the determinant* $| R_{ik} |$ *is different from zero on* x_1x_2, *the strengthened condition of Legendre holds.*

For, suppose there exists a point x_3 on x_1x_2 and a set of constants (π) such that $R_{ik}\pi_i\pi_k < 0$. The inequality

$$2\omega(x, \rho\pi, \pi) < 0 \tag{3.7}$$

accordingly holds at $x = x_3$ for $\rho = 0$ and hence also on an interval $x'x''$ for all values of ρ which are in absolute value less than a small constant ρ_0. Divide the interval $x'x''$ into m (m arbitrary) successive intervals h_α $(\alpha = 1, \cdots, m)$ of equal length and denote the center and the length of h_α by t_α, $2l$ respectively. Choose the constant δ as the smaller of ρ_0 and l and set

$$\eta_{i\alpha} = (x - t_\alpha + \delta)\pi_i \qquad (t_\alpha - \delta \leqq x \leqq t_\alpha),$$

$$\eta_{i\alpha} = (-x + t_\alpha + \delta)\pi_i \qquad (t_\alpha \leqq x \leqq t_\alpha + \delta),$$

and $\eta_{i\alpha} \equiv 0$ elsewhere. From the inequality (3.7) it follows readily that

$$J(\eta_\alpha) = J(\eta_\alpha, \eta_\alpha) = \int_{t_\alpha-\delta}^{t_\alpha+\delta} 2\omega(x, \eta_\alpha, \eta_\alpha')\, dx < 0\,.$$

Moreover for $\alpha \neq \beta$ the expression $\Omega(\eta_\alpha, \eta_\beta)$ is identically zero and hence $J(\eta_\alpha, \eta_\beta) = 0$. It follows from Theorem 3.4 that in this case the type number of J is at least m and hence infinite since m is arbitrary. This proves the theorem.

We shall show in §5 below that if the strengthened condition of Legendre holds, the type number m of J is finite and is equal to the sum of the orders of the conjugate points of x_1 between x_1 and x_2.

4. **Two lemmas.** In this section we shall establish two lemmas which will be useful in proving the existence of a minimal set. For this purpose we shall assume that the strengthened condition of Legendre holds.

It is well known that for two accessory extremals η_i, u_i the expression

$$\eta_i\omega_{u'_i} - u_i\omega_{\eta'_i} \tag{4.1}$$

is a constant, where the arguments in $\omega_{\eta'_i}$, $\omega_{u'_i}$ are respectively η_i, η'_i and u_i, u'_i. This fact follows at once from the relations

$$\Omega(\eta, u) = \Omega(u, \eta), \qquad \Omega(\eta, u) = \eta_i L_i(u) + (d/dx)(\eta_i\, \omega_{u'_i}).$$

If the constant (4.1) is zero, the extremals are said to be *conjugate* to each other. A set of linearly independent mutually conjugate extremals is said to form a *conjugate system*. It is clear that a set of n linearly independent accessory extremals $\eta_{ik}(x)$ having $\eta_{ik}(x_1) = 0$ forms a conjugate system.

LEMMA 4.1. *If an admissible arc η_i and a conjugate system η_{ik} are such that the matrices*

$$\|\, \eta_{ik} \;\; \eta_i \,\|, \qquad \|\, \eta_{ik} \,\| \qquad (i, k = 1, \cdots, n) \tag{4.2}$$

have the same rank at each point on x_1x_2, then the functions $\eta_i(x)$ are expressible in the form

$$\eta_i = \eta_{ik}a_k(x). \tag{4.3}$$

The functions $a_k(x)$ are of class D' except possibly at the zeros of the determinant

$$D(x) = |\, \eta_{ik}(x) \,|.$$

At a zero x_3 of $D(x)$ the functions a_k are continuous and the functions $\eta_{ik}a'_k$ have unique right and left hand limits.

To prove this, we note that for values of x at which the determinant $D(x)$ is different from zero the solutions $a_k(x)$ of equations (4.3) are of the form

$$a_j = D_j/D,$$

where the functions $D_j(x)$ are obtained from D by replacing η_{ij} by η_i. Consider now a value x_3 at which $D(x_3)$ has rank $n - r$. We may suppose that the functions η_{ik} have been chosen so that the first r columns of $D(x_3)$ are composed of zeros. Moreover, we may assume that $\eta_i(x_3) = 0$ since this can be brought about by adding to η_i a suitable linear combination of the extremals η_{ik}, in view of the fact that the matrices (4.2) have the same rank. By the law of the mean with integral remainder we see that

$$\eta_i = (x - x_3)W_i, \qquad \eta_{ij} = (x - x_3)W_{ij} \qquad (j = 1, \cdots, r),$$

where $W_i(x)$, $W_{ij}(x)$ are continuous in x and have

$$W_i(x_3) = \eta'_i(x_3), \qquad W_{ij}(x_3) = \eta'_{ij}(x_3).$$

It follows that in a neighborhood of $x = x_3$ we have

$$D = (x - x_3)^r\Delta(x), \quad D_k = (x - x_3)^r\Delta_k(x) \qquad (k = 1, \cdots, n),$$

where Δ, Δ_k are continuous in x and $\Delta(x_3) \neq 0$, as will be shown below. If we set

$$a_k(x_3) = \Delta_k(x_3)/\Delta(x_3),$$

it is clear that the functions $a_k(x)$ will be continuous at $x = x_3$. Moreover from the equations

$$\eta_i' = \eta_{ik}' a_k + \eta_{ik} a_k'$$

it follows that the expressions $\eta_{ik} a_k'$ have unique right and left hand limits at $x = x_3$, as was to be proved.

It remains to show that the determinant

$$\Delta(x_3) = | \eta_{ij}' \ \eta_{ih} | \qquad (j = 1, \cdots, r; h = r + 1, \cdots, n)$$

is different from zero. If $\Delta(x_3)$ were zero, there would exist constants a_j, b_h not all zero such that

$$\eta_{ij}' a_j = \eta_{ih} b_h$$

at $x = x_3$. Since the extremals $\eta_i = \eta_{ij} a_j$, $u_i = \eta_{ih} b_h$ are mutually conjugate and $\eta_i(x_3) = 0$, the expression (4.1) would reduce at $x = x_3$ to the form

$$0 = R_{ik} \eta_i' u_k = R_{ik} u_i u_k.$$

This is possible only in case $b_k = 0$ by virtue of the Legendre condition. But in this case we would also have $\eta_i(x_3) = \eta_i'(x_3) = 0$ and hence $a_j = 0$, since the solutions η_{ij} are linearly independent. It follows that $\Delta(x_3) \neq 0$, and Lemma 4.1 is established.

Consider now the conjugate system $\eta_{ik}(x)$ and let

$$u_i = \eta_{ik} a_k(x), \qquad v_i = \eta_{ik} a_k'(x),$$

where the functions $a_k(x)$ are of class C' and a_k' are their derivatives. If we denote by primes differentiation with respect to x when a_k, a_k' are considered as constants, we have the following immediate relations

$$u_i' = \eta_{ik}' a_k, \qquad v_i' = \eta_{ik}' a_k',$$

$$(\omega_{u_i'})' = \omega_{u_i}, \qquad (\omega_{v_i'})' = \omega_{v_i}, \qquad u_i \omega_{v_i'} - v_i \omega_{u_i'} = 0,$$

$$(d/dx)\omega_{u_i'} = (\omega_{u_i'})' + \omega_{v_i'} = \omega_{u_i} + \omega_{v_i'}.$$

If now we set

$$\eta_i(x) = u_i(x), \qquad \eta_i'(x) = u_i'(x) + v_i(x) = d\eta_i/dx,$$

it is readily seen that[7]

$$\begin{aligned} 2\omega(\eta, \eta') &= 2\omega(u, u' + v) \\ &= 2\omega(u, u') + 2v_i\omega_{u'_i} + R_{ik}v_iv_k \\ &= u_i\omega_{u_i} + u'_i\omega_{u'_i} + 2v_i\omega_{u'_i} + R_{ik}v_iv_k \\ &= u_i(\omega_{u_i} + \omega_{v'_i}) + (u'_i + v_i)\omega_{u'_i} + R_{ik}v_iv_k \\ &= (d/dx)(\eta_i\omega_{u'_i}) + R_{ik}v_iv_k. \end{aligned}$$

It follows that if η_i is an admissible variation, we have

$$J(\eta) = \int_{x_1}^{x_2} R_{ik}v_iv_k\, dx \geqq 0\,. \tag{4.4}$$

The equality holds only in case the functions v_i are identically zero, that is, only in case $a'_k \equiv 0$ and η_i is an extremal arc. The relation (4.4) holds even if the functions a_k are not of class C'. It is sufficient that the functions a_k have the properties described in Lemma 4.1, as one readily verifies. Hence we have the following obvious consequence:

LEMMA 4.2. *Suppose the strengthened condition of Legendre holds on x_1x_2. If $\eta_i(x)$ is a set of admissible variations such that the matrices* (4.2) *have the same rank at each point of x_1x_2, then $J(\eta) \geqq 0$, the equality holding only in case η_i is an accessory extremal.*

An analogous lemma has been used by Morse[8] for the problem of Lagrange. We have the following

COROLLARY. *Suppose the strengthened condition of Legendre holds on x_1x_2. If there are no conjugate points of x_1 on x_1x_2, every accessory extremal u_i affords a minimum to J relative to arcs η_i of class D' joining its end points.*

For if the arc η_i joins the end points of the extremal u_i, then we have $J(\eta - u) \geqq 0$, by Lemma 4.2. Moreover $J(\eta, u) = J(u, u)$, as can be seen by integration by parts. Hence

$$J(\eta - u) = J(\eta) - 2J(\eta, u) + J(u) = J(\eta) - J(u) \geqq 0\,,$$

as was to be proved.

5. The existence of minimal sets.

We begin with the following

LEMMA 5.1. *Suppose the strengthened condition of Legendre holds and let x_3*

[7] The arguments here used are due to Bliss, *The transformation of Clebsch in the Calculus of Variations*, Proceedings of the International Mathematical Congress, Toronto, vol. 1 (1924), pp. 589–603. See also Bliss, *The problem of Lagrange in the Calculus of Variations*, American Journal of Mathematics, vol. 52 (1930), pp. 738–9. Further references are given by Bliss.

[8] Morse, *Sufficient conditions in the problem of Lagrange without assumptions of normalcy*, Trans. Amer. Math. Soc., vol. 37 (1935).

be a conjugate point of x_1 of order r. For every interval $x'x''$ sufficiently small with x_3 in its interior there exists a set of r admissible variations $\xi_{ij}(x)$ $(j = 1, \cdots, r)$ of class C'' having $\xi_{ij} \equiv 0$ on the interval $x''x_2$ and

$$J(\xi_h, \xi_j)\, a_h a_j < 0 \qquad (h, j = 1, \cdots, r) \tag{5.1}$$

for every set of constants $(a) \neq (0)$. Moreover, on the interval x_1x' the arc ξ_{ij} defines an accessory extremal and at $x = x_3$ the matrix

$$\| \eta_{ik}\ \xi_{ij} \| \qquad (i, k = 1, \cdots, n; j = 1, \cdots, r) \tag{5.2}$$

has rank n, where η_{ik} is a set of n linearly independent accessory extremals having $\eta_{ik}(x_1) = 0$.

In order to obtain this result we suppose that the system η_{ik} has been chosen so that $\eta_{ij}(x_3) = 0$ $(j = 1, \cdots, r)$ and suppose that the interval $x'x''$ has been chosen so that there are no pairs of conjugate points on $x'x''$. Let u_{ij} be a set of r accessory extremals having

$$u_{ij}(x') = \eta_{ij}(x'), \qquad u_{ij}(x'') = 0,$$

and let ξ_{ij} be identical with η_{ij} on x_1x', identical with u_{ij} on $x'x''$, and identically zero on $x''x_2$. The functions ξ_{ij} have all the properties described in the lemma except that of being of class C''. In order to prove this, let a_j be a set of r constants not all zero and let $\xi_i = \xi_{ij}\, a_j$. If we set $\eta_i = \eta_{ij}\, a_j$ on x_1x_3 and $\eta_i \equiv 0$ on x_3x_2, then $\eta_i(x') = \xi_i(x')$, $\eta_i(x'') = \xi_i(x'')$ and

$$J(\xi) - J(\eta) = \int_{x'}^{x''} 2\omega(\xi, \xi')\, dx - \int_{x'}^{x''} 2\omega(\eta, \eta')\, dx .$$

On the interval $x'x''$ the arc ξ_i is an extremal arc whose end points are joined by the arc η_i. Hence we have $J(\xi) < J(\eta)$, by the corollary to Lemma 4.2, since the strengthened condition of Legendre holds and the interval $x'x''$ contains no pairs of conjugate points. Moreover by the usual integration by parts it is seen that

$$J(\eta) = \int_{x_1}^{x_3} 2\omega(\eta, \eta')\, dx = \int_{x_1}^{x_3} \Omega(\eta, \eta)\, dx = \eta_i \omega_{\eta_i'} \Big|_{x_1}^{x_3} = 0 .$$

It follows that $J(\xi) < 0$, and the inequality (5.1) is established.

In order to show that the matrix (5.2) has rank n at $x = x_3$ we note first that by the usual integration by parts we have

$$\begin{aligned} J(\xi) &= \int_{x_1}^{x'} \Omega(\xi, \xi)\, dx + \int_{x'}^{x''} \Omega(\xi, \xi)\, dx \\ &= \xi_i(x_1 - 0)\omega_{\xi_i'}(x_1 - 0) - \xi_i(x_1 + 0)\omega_{\xi_i'}(x_1 + 0) \qquad (5.3)\\ &= \xi_i(x_1 + 0)\omega_{\xi_i'}(x_1 - 0) - \xi_i(x_1 - 0)\omega_{\xi_i'}(x_1 + 0) \\ &= (\zeta_{ij}\, u_{ih} - \eta_{ij} v_{ih})\, a_j\, a_h \qquad (h, j = 1, \cdots, r), \end{aligned}$$

where ζ_{ik}, v_{ih} denote the values of $\omega_{\eta_i'}$ along η_{ik} and u_{ih} respectively. It is clear that the matrix (5.2) has rank n at $x = x_3$ if and only if the determinant

$$| u_{ij} \quad \eta_{il} | \qquad (j = 1, \cdots, r; l = r + 1, \cdots, n)$$

is different from zero. If this determinant were zero at $x = x_3$, there would exist constants a_j, b_l not all zero such that at $x = x_3$

$$u_{ij} a_j = \eta_{il} b_l .$$

Since $\eta_{ij}(x_3) = 0$ $(j = 1, \cdots, r)$ and the system η_{ik} is a conjugate system, it would follow that at $x = x_3$

$$\begin{aligned} 0 &= (\zeta_{ij}\eta_{il} - \eta_{ij}\zeta_{il})b_l = \zeta_{ij}\eta_{il}b_l \\ &= \zeta_{ij}u_{ih}\, a_h = (\zeta_{ij}u_{ih} - \eta_{ij}v_{ih})a_h . \end{aligned}$$

This is possible only in case $a_h = 0$ in view of the relation (5.3) and $J(\xi) < 0$ for $(a) \neq (0)$. But in this case we should have $b_l = 0$ also, since x_3 is a conjugate point of order r. It follows that the matrix (5.2) has rank n at $x = x_3$. It is clear that the corners of ξ_{ij} can be rounded off so as to obtain functions ξ_{ij} of class C'' having the properties just described. This completes the proof of Lemma 5.1.

LEMMA 5.2. *Let s be the sum of the orders of the conjugate points of x_1 between x_1 and x_2. If the strengthened condition of Legendre holds, there exists a set of s admissible variations $\xi_{i\alpha}$ $(\alpha = 1, \cdots, s)$ of class C'' which have*

$$J(\xi_\alpha, \xi_\beta)a_\alpha a_\beta < 0 \qquad (\alpha, \beta = 1, \cdots, s) \tag{5.4}$$

for every set of constants $(a) \neq (0)$ and which are such that for every set of admissible variations $\eta_i(x)$ there is a set of constants μ_β such that the matrices

$$\| \eta_{ik} \quad \eta_i + \mu_\beta \xi_{i\beta} \|, \qquad \| \eta_{ik} \| \tag{5.5}$$

have the same rank at each point on x_1x_2, where η_{ik} is a conjugate system having $\eta_{ik}(x_1) = 0$.

For let conjugate points of x_1 between x_1 and x_2 be given by the sequence

$$t_1 < t_2 < \cdots < t_q$$

and let r_h be the order of the conjugate point t_h. Associate with each conjugate point t_h an interval $x_h'x_h''$ as in Lemma 5.1 and such that these intervals have no points in common. Let $s_h = r_1 + \cdots + r_h$ and let $\xi_{i\gamma}$ $(\gamma = s_{h-1} + 1, \cdots, s_h)$ be a set of admissible variations related to t_h as in Lemma 5.1. The functions $\xi_{i\beta}$ so defined can be shown to have the properties described in Lemma 5.2. In view of the inequality (5.1) it is clear that the condition (5.4) will be established if we show that

$$J(\xi_\alpha, \xi_\beta) = 0$$

whenever $\xi_{i\alpha}$, $\xi_{i\beta}$ belong to different conjugate points t_h, t_j. We may suppose $j > h$. On the interval $x_1x'_j$ the arc $\xi_{i\beta}$ will be an extremal arc. Moreover $\xi_{i\alpha}$ is identically zero on the interval x'_jx_2. It follows that

$$J(\xi_\alpha, \xi_\beta) = \int_{x_1}^{x'_j} \Omega(\xi_\alpha, \xi_\beta)dx = \xi_{i\alpha}\omega_{\eta'_i}(\xi_\beta, \xi'_\beta)\Big|_{x_1}^{x'_j} = 0,$$

as was to be proved.

In order to prove the last part of Lemma 5.2, it should be noted that at $x = t_h$ the functions $\xi_{i\tau}$ $(\tau = 1, \cdots, s_{h-1})$ are all zero. It follows that the matrices (5.5) involve only the last $s - s_{h-1}$ constants μ_β at $x = t_h$. We may accordingly so choose the last $s - s_{q-1}$ constants μ_β that the matrices (5.5) have the same rank at $x = t_q$. The constants μ_γ $(\gamma = s_{q-2} + 1, \cdots, s_{q-1})$ are then chosen so that the matrices (5.5) have the same rank at $x = t_{q-1}$. Proceeding in this manner, we finally obtain a set of constants μ_β such that these matrices have the same rank at each point on x_1x_2. This completes the proof of Lemma 5.2.

We can now establish the following important result:

THEOREM 5.1. *Let the strengthened condition of Legendre hold and let s be the sum of the orders of the conjugate points of x_1 between x_1 and x_2. Then there exists a set of s natural isoperimetric conditions*

$$J(\xi_\beta, \eta) = 0 \qquad (\beta = 1, \cdots, s)$$

such that $J(\eta) \geqq 0$ for every admissible arc η_i satisfying these conditions, the equality holding only in case η_i is an accessory extremal. Moreover these conditions are such that the relation (5.4) *holds for every set* $(a) \neq (0)$.

For let $\xi_{i\beta}$ $(\beta = 1, \cdots, s)$ be a set of s functions having the properties described in Lemma 5.2, and consider the set of natural isoperimetric conditions

$$J(\xi_\beta, \eta) = 0 \qquad (\beta = 1, \cdots, s). \tag{5.6}$$

Let $\eta_i(x)$ be an admissible variation satisfying these conditions and let μ_β be a set of constants such that the matrices (5.5) have the same rank at each point on x_1x_2. From Lemma 4.2 it follows that

$$J(\eta + \mu_\beta\xi_\beta) \geqq 0,$$

the equality holding only in case the arc $\eta_i + \mu_\beta\xi_{i\beta}$ is an extremal arc. By the use of the inequality (5.4) and the relation

$$\begin{aligned} J(\eta + \mu_\beta\xi_\beta) &= J(\eta) + 2\mu_\beta J(\xi_\beta, \eta) + J(\mu_\beta\xi_\beta) \\ &= J(\eta) + J(\xi_\alpha, \xi_\beta)\mu_\alpha\mu_\beta \geqq 0, \end{aligned}$$

it is found that $J(\eta) \geqq 0$, the equality holding only in case η_i is an extremal arc. The set (5.6) is accordingly a minimal set by virtue of the corollary to Theorem 3.1. This proves Theorem 5.1.

COROLLARY 1. *If the strengthened condition of Legendre holds, the type number of J is equal to the sum of the conjugate points of x_1 between x_1 and x_2.*

The following corollary will be useful in §9 below:

COROLLARY 2. *Suppose the strengthened condition of Legendre holds and there is no conjugate point of x_1 on the interval $x_3 \leqq x < x_2$. If $\eta_i(x)$ is a broken extremal whose corner is at $x = x_3$ and which has $\eta_i(x_1) = \eta_i(x_2) = 0$, then $J(\eta) \geqq 0$, the equality holding only in case η_i is an accessory extremal.*

This result follows because the arc η_i satisfies the conditions (5.6), as one readily verifies by the usual integration by parts.

6. Oscillation, separation, and comparison theorems. The results described in §§3 and 5 enable us to state the following new oscillation theorem. It is assumed throughout this section that the strengthened condition of Legendre holds on x_1x_2. Moreover, a conjugate point is counted a number of times equal to its order.

THEOREM 6.1. *There are (at least) m conjugate points of x_1 between x_1 and x_2 if and only if there exists a set of m admissible variations $\eta_{i\beta}$ such that the inequality*

$$J(\xi_\alpha, \xi_\beta)a_\alpha a_\beta < 0 \qquad (\alpha, \beta = 1, \cdots, m) \tag{6.1}$$

holds for every set of constants $(a) \neq (0)$. If the variations $\eta_{i\beta}$ form a maximal set of admissible variations satisfying the condition (6.1), *then there are exactly m conjugate points of x_1 between x_1 and x_2.*

This result follows at once from Theorem 3.4 and Corollary 1 of Theorem 5.1.

COROLLARY. *The number of conjugate points of x_1 on x_1x_2 is equal to the number of conjugate points of x_2 on x_1x_2.*

The following generalization of the Sturm separation theorem, first proved by Morse[9] and later by Hu,[10] is immediate.

THEOREM 6.2. *If $x_0 > x_1$, then the k^{th} conjugate point of x_1 precedes the k^{th} conjugate of x_0.*

This result follows at once from the above corollary since if x_2 is the k^{th} conjugate point of x_1, there must be fewer than k conjugate points of x_2 (and hence of x_0) on the interval x_0x_2.

In a similar manner one can prove the stronger theorem:

THEOREM 6.3. *The k^{th} conjugate point of x_1 varies continuously with x_1 and advances or regresses with it.*

Theorem 6.1 also gives us at once the usual comparison theorems. In the following theorem the integral $J^*(\eta)$ is assumed to be of the same type as $J(\eta)$.

THEOREM 6.4. *Let m, m^* be the sums of the orders of the conjugate points of x_1 between x_1 and x_2 and let d, d^* be the orders of the point x_2 as a conjugate point*

[9] Morse, *A generalization of the Sturm separation and comparison theorems in n-space*, Mathematische Annalen 103 (1930), pp. 52–69.

[10] Hu, *The problem of Bolza and its accessory boundary value problem*, Contributions to the Calculus of Variations 1931–32, The University of Chicago Press, pp. 361–444.

of x_1 relative to the functionals J, J^* respectively. If $J(\eta) \geqq J^*(\eta)$ for every set of admissible variations $(\eta) \not\equiv (0)$, then

$$m^* \geqq m, \qquad m^* + d^* \geqq m + d. \tag{6.2}$$

Moreover if $J(\eta) > J^*(\eta)$ for all such variations, then

$$m^* \geqq m + d. \tag{6.3}$$

To prove this, let

$$J(\xi_\alpha, \eta) = 0 \qquad (\alpha = 1, \cdots, m)$$

be a minimal set for $J(\eta)$ and let $\xi_{i\gamma}$ $(\gamma = m + 1, \cdots, m + d)$ be a maximal set of linearly independent accessory extremals vanishing at x_1 and x_2. We have accordingly

$$J(\xi_\alpha, \xi_\beta)a_\alpha a_\beta \leqq 0 \qquad (\alpha, \beta = 1, \cdots, m + d),$$

the equality holding only in case the first m a's are zero. It follows that

$$J^*(\xi_\alpha, \xi_\beta)a_\alpha a_\beta \leqq 0 \qquad (\alpha, \beta = 1, \cdots, m + d), \tag{6.4}$$

where the equality sign can be excluded if the first m a's are not all zero. Hence $m^* \geqq m$. Moreover if $J(\eta) > J^*(\eta)$ for every set of admissible variations $(\eta) \not\equiv (0)$, then the condition (6.4) holds for every set $(a) \neq (0)$ with the equality sign excluded. Hence in this case $m^* \geqq m + d$. In order to prove the inequality $m^* + d^* \geqq m + d$, we may suppose that

$$J^*(\xi_\alpha, \xi_\beta) = 0 \qquad (\alpha \neq \beta).$$

By the process used in the proof of Theorem 3.2 we can replace a variation $\xi_{i\beta}$ which is not an extremal for J^* and which has $J^*(\xi_\beta) = 0$ by one for which $J^*(\xi_\beta) < 0$. The number of extremals in the set $\xi_{i\alpha}$ cannot exceed d^*. If these are deleted, the remaining functions will form a subset of a set of functions defining a minimal set for J^*. Hence $m^* + d^* \geqq m + d$, as was to be proved.

II

The accessory problem in the variable end point case

In §§7–16 we shall study the functional

$$J(\eta) = 2q(\eta_1, \eta_2) + \int_{x_1}^{x_2} 2\omega(x, \eta, \eta')\, dx$$

subject to a set of end conditions

$$\Psi_\mu(\eta) = a_{\mu i}\eta_i(x_1) + b_{\mu i}\eta_i(x_2) = 0 \qquad (\mu = 1, \cdots, p \leqq 2n),$$

where $2q$ is a symmetric quadratic form in η_{i1} and η_{i2}, and 2ω is a quadratic form in η_i, η_i' of the type described above. This problem arises in the study of the second variation of the general variable end point problems in parametric and non-parametric form.

In these sections we define natural isoperimetric conditions for the functional $J(\eta)$ and show how they can be used to characterize the extremals of $J(\eta)$ satisfying the conditions $\Psi_\mu = 0$ and the associated transversality conditions. This characterization leads to a new approach to oscillation and comparison theorems.

Besides the assumptions made in §1 we shall assume that the matrix

$$\| \Psi_{\mu\eta_{i1}} \quad \Psi_{\mu\eta_{i2}} \|$$

has rank p. An arc η_i of class D' satisfying the conditions $\Psi_\mu = 0$ will be called an *admissible arc* in the variable end point case.

7. Orders of degeneracy and concavity. The Euler equations and the transversality conditions of $J(\eta)$ subject to the conditions $\Psi_\mu = 0$ are the equations

$$L_i(\eta) = \omega_{\eta_i} - (d/dx)\omega_{\eta_i'} = 0\,, \tag{7.1}$$

$$T_{i1}(\eta, l) = q_{\eta_{i1}} + l_\mu \Psi_{\mu\eta_{i1}} - \omega_{\eta_i'}(x_1) = 0\,,$$

$$T_{i2}(\eta, l) = q_{\eta_{i2}} + l_\mu \Psi_{\mu\eta_{i2}} + \omega_{\eta_i'}(x_2) = 0\,. \tag{7.2}$$

The equations (7.2) will be termed *accessory transversality conditions* and the solutions η_i of equations (7.1) of class C'' will be called *accessory extremals.* One readily verifies that $J(\eta) = 0$ along an accessory extremal which satisfies the conditions (7.2) and $\Psi_\mu = 0$. The maximum number d of linearly independent accessory extremals satisfying the conditions (7.2) and $\Psi_\mu = 0$ will be called the *order of degeneracy* of $J(\eta)$ relative to the conditions $\Psi_\mu = 0$.

Let h be the maximum number of linearly independent accessory extremals $\eta_{i\beta}(\beta = 1, \cdots, r)$ satisfying the conditions $\Psi_\mu = 0$ and having $J(\eta_\beta a_\beta) < 0$ for every set of constants $(a) \neq (0)$. Let k be the maximum number of linearly independent accessory extremals $\eta_{i\gamma}(\gamma = h + 1, \cdots, h + k)$ which have $\eta_{i\gamma}(x_1) = \eta_{i\gamma}(x_2) = 0$ and which are such that no proper linear combination of these extremals satisfies the conditions (7.2). The number $r = h + k$ will be called the *order of concavity* of $J(\eta)$ relative to the conditions $\Psi_\mu = 0$. It is clear that $r + d \leqq 2n - p + d$ and hence that $r \leqq 2n - p$, where p is the number of conditions $\Psi_\mu = 0$.

For the periodic case, the number h has been called the order of concavity of $J(\eta)$ by Morse. We prefer the definition here given since it measures the difference between the type numbers of $J(\eta)$ in the fixed and variable end point cases, as will be seen in §10 below.

Let

$$(7.3)\qquad \begin{aligned} Q(\xi, \eta) &= \xi_{i1} q_{\eta_{i1}} + \xi_{i2} q_{\eta_{i2}}, \\ \Omega(\xi, \eta) &= \xi_i \omega_{\eta_i} + \xi'_i \omega_{\eta'_i}, \\ J(\xi, \eta) &= Q(\xi, \eta) + \int_{x_1}^{x_2} \Omega(\xi, \eta)\, dx. \end{aligned}$$

With these notations in mind we can prove the following result:

LEMMA 7.1. *Suppose that the point x_2 is not conjugate to x_1 and that the determinant $|R_{ik}|$ is different from zero on x_1x_2. If $\xi_{i\beta}$ $(\beta = 1, \cdots, r)$ forms a maximal set of accessory extremals satisfying the conditions $\Psi_\mu = 0$ and having*

$$J(\xi_\alpha, \xi_\beta) a_\alpha a_\beta < 0 \qquad (\alpha, \beta = 1, \cdots, r)$$

for every set of constants $(a) \neq (0)$, then every accessory extremal η_i satisfying the conditions

$$\Psi_\mu = 0, \qquad J(\xi_\alpha, \eta) = 0$$

and having $J(\eta) = 0$ also satisfies the accessory transversality conditions (7.2).

To prove this, we first note that if η_i has the properties described in the theorem, then the determinant

$$(7.4)\qquad \begin{vmatrix} J(\eta, \eta_\nu) \\ J(\xi_\alpha, \eta_\nu) \\ \Psi_\mu(\eta_\nu) \end{vmatrix} \qquad (\nu = 1, \cdots, r + p + 1)$$

must be zero for every set of $r + p + 1$ accessory extremals $\eta_{i\nu}$. Otherwise for some set $\eta_{i\nu}$ the equations

$$J(\eta_\nu, \eta_\nu)\, b_\nu = 1, \qquad J(\xi_\alpha, \eta_\nu) b_\nu = 0, \qquad \Psi_\mu(\eta_\nu) b_\nu = 0$$

would have a solution $(b) \neq (0)$.[11] Hence by setting $\bar\eta_i = \eta_{i\nu} b_\nu$ we should have

$$J(\eta - c\bar\eta) = J(\eta) - 2cJ(\eta, \bar\eta) + c^2 J(\bar\eta) = -2c + c^2 J(\bar\eta),$$
$$J(\xi_\alpha, \eta - c\bar\eta) = J(\xi_\alpha, \eta) - cJ(\xi_\alpha, \bar\eta) = 0.$$

For c sufficiently small and positive we should have $J(\eta - c\bar\eta) < 0$, contrary to our choice of the functions $\xi_{i\beta}$. It follows that the determinant (7.4) is always zero and hence that there exists a set of constants l_0, h_α, l_μ not all zero such that the equation

$$l_0 J(\eta, u) + h_\alpha J(\xi_\alpha, u) + l_\mu \Psi_\mu(u) = 0$$

holds for every accessory extremal u_i. Substituting the functions $\xi_{i\beta}$ for u_i in this last expression gives $h_\alpha = 0$ $(\alpha = 1, \cdots, r)$. If now we replace u_i

[11] See Bliss, *The problem of Lagrange in the Calculus of Variations*, American Journal of Mathematics, vol. 52 (1930), p. 682.

by a set of p accessory extremals, no proper linear combination of which satisfies the equations $\Psi_\mu = 0$, it is seen that $l_0 \neq 0$, since otherwise the constants l_0, l_μ would be all zero. This is not the case. We may accordingly suppose $l_0 = 1$. Moreover, every pair of points $u_i(x_1)$, $u_i(x_2)$ can be joined by an extremal u_i, since the point x_2 is not conjugate to x_1. It follows that the expression

$$J(\eta, u) + l_\mu \Psi_\mu(u) = T_{i1}(\eta, l)u_i(x_1) + T_{i2}(\eta, l)u_i(x_2) = 0$$

must be an identity in the values $u_i(x_1)$, $u_i(x_2)$ and hence that η_i satisfies the conditions (7.2) with the constants l_μ, as was to be proved.

From this lemma we obtain at once the following method of determining the orders of concavity and degeneracy of J in case the point x_2 is not conjugate to x_1.

THEOREM 7.1. *Suppose that the point x_2 is not conjugate to x_1 and that $|R_{ik}| \neq 0$ on x_1x_2. Let $\eta_{i\alpha}$ $(\alpha = 1, \cdots, 2n - p)$ be a set of $2n - p$ linearly independent accessory extremals satisfying the conditions $\Psi_\mu = 0$. The order of concavity and the order of degeneracy of J are equal respectively to the negative type number and the nullity of the quadratic form*

$$H(z) = J(\eta_\alpha, \eta_\beta)z_\alpha z_\beta \qquad (\alpha, \beta = 1, \cdots, 2n - p).$$

The quadratic form $H(z)$ is the classical one used in the study of variable end point problems. It was first used by A. Mayer[12] and has been used in various forms by several writers. Its importance has been emphasized recently by Bliss.[13]

8. **Natural isoperimetric conditions.** A condition of the type

$$J(\xi, \eta) = Q(\xi, \eta) + \int_{x_1}^{x_2} \Omega(\xi, \eta)\, dx = 0,$$

where $\xi_i(x)$ is a function of class C'' satisfying the conditions $\Psi_\mu = 0$ will be called a *natural isoperimetric condition. It is satisfied by every accessory extremal satisfying the conditions $\Psi_\mu = 0$ and the accessory transversality conditions* (7.2), as one readily verifies by substitution and the usual integration by parts.

Consider now a set of m natural isoperimetric conditions

$$J(\xi_\alpha, \eta) = 0 \qquad (\alpha = 1, \cdots, m). \tag{8.1}$$

The Euler equations and the transversality conditions of J relative to the conditions (8.1) and the end conditions $\Psi_\mu = 0$ are the equations

$$L_i(\eta + \mu_\beta \xi_\beta) = 0, \qquad \mu_\beta' = 0, \tag{8.2}$$

$$T_{i1}(\eta + \mu_\beta \xi_\beta, l) = 0, \qquad T_{i2}(\eta + \mu_\beta \xi_\beta, l) = 0, \tag{8.3}$$

[12] A. Mayer, *Zur Aufstellung der Kriterien des Maximums und Minimums der einfachen Integrale bei variablen Grenzwerten*, Leipziger Berichte, vol. 36 (1884), pp. 99–128.

[13] Bliss, *The problem of Bolza in the Calculus of Variations*, Annals of Mathematics, (2), vol. 33 (1932), pp. 261–274.

where L_i, T_{i1}, T_{i2} are the functions defined by equations (7.1) and (7.2). A solution η_i, μ_β of equations (8.2) of class C'' will be called an *isoperimetric extremal.*

A set of natural isoperimetric conditions (8.1) will be called a *proper set* if the determinant

$$| J(\xi_\alpha, \xi_\beta) |$$

is different from zero. We have the following analogue of Theorem 2.1, which can be proved by the same methods.

THEOREM 8.1. *If the set* (8.1) *is a proper set, then the multipliers μ_β belonging to an isoperimetric extremal η_i, μ_β satisfying the conditions* (8.1), (8.3), *and $\Psi_\mu = 0$ are all zero.*

The analogues of Theorem 2.2 and its corollaries can also be established by the arguments used in §2. The following theorem is the analogue of the second corollary.

THEOREM 8.2. *If the determinant $| R_{ik} |$ is different from zero, then an admissible arc η_i which satisfies all natural isoperimetric conditions is necessarily an extremal arc satisfying the transversality conditions* (7.2).

The results described in §3 can be extended at once to the case here considered. The proofs of the following theorems are identical with those given in §3.

THEOREM 8.3. *If the set* (8.1) *forms a minimal set of natural isoperimetric conditions, there exists a proper minimal set composed of the same number of conditions.*

THEOREM 8.4. *If the set* (8.1) *is a proper minimal set of natural isoperimetric conditions, the inequality*

$$J(\xi_\alpha, \xi_\beta)a_\alpha a_\beta < 0 \qquad (\alpha, \beta = 1, \cdots, m) \tag{8.4}$$

is true for every set of constants $(a) \neq (0)$. *Conversely, if the set* (8.1) *is a maximal set of natural isoperimetric conditions satisfying the condition* (8.4), *it is a proper minimal set.*

THEOREM 8.5. *The number of conditions in a minimal set is always the same.*

The number of conditions in a minimal set will be called the *type number* of the functional $J(\eta)$ relative to the conditions $\Psi_\mu = 0$. If there exists no minimal set, the type number of J will be said to be infinite. In view of Theorem 8.5 the type number of J is a well defined concept.

THEOREM 8.6. *If the functions $\eta_{i\beta}$ $(\beta = 1, \cdots, m)$ form a set of m admissible variations having*

$$J(\xi_\alpha, \xi_\beta)a_\alpha a_\beta < 0 \qquad (\alpha, \beta = 1, \cdots, m) \tag{8.5}$$

for every set of constants $(a) \neq (0)$, *the type number of J is at least equal to m. If the set $\eta_{i\beta}$ forms a maximal set of such variations, the type number of J is equal to m.*

This follows from Theorem 8.4 since it is clear that the variations $\eta_{i\beta}$ can

be replaced by admissible variations $\xi_{i\beta}$ of class C'' without disturbing the inequality (8.5).

We have the following interesting corollary which will be useful in the proof of Theorem 12.6 below.

COROLLARY. *If $\omega(\eta, \eta') \equiv 0$, the type number of J is equal to the negative type number of the quadratic form $2q(\eta_1, \eta_2)$ subject to the auxiliary conditions $\Psi_\mu = 0$.*

The condition of Legendre has been defined in §3. We have the following theorem concerning this condition.

THEOREM 8.7. *If the type number m of J is finite, then the condition of Legendre holds on x_1x_2. Moreover, if the determinant $| R_{ik} |$ is different from zero on x_1x_2, the strengthened condition of Legendre holds on x_1x_2.*

This result follows at once from Theorem 3.5 since the type number of J relative to the conditions $\eta_i(x_1) = \eta_i(x_2) = 0$ is clearly less than m.

9. **An equivalent problem.** In many cases our arguments can be simplified a great deal if we can assume without loss of generality that the point x_2 is not conjugate to x_1. In this section we shall show that this can be done. We do so by showing that we can always replace our problem by a second problem which has the same type number and order of degeneracy and for which the point x_2 is not conjugate to x_1. In fact we can choose this new problem so as to have the same order of concavity and the same number of conjugate points of x_1 between x_1 and x_2 as our original problem. We assume that the strengthened condition of Legendre holds.

Suppose now that we have given the functional

$$J(\eta) = 2q(\eta_1, \eta_2) + \int_{x_1}^{x_2} 2\omega(x, \eta, \eta')\, dx$$

subject to the end conditions

$$\Psi_\mu(\eta) = a_{\mu i}\eta_i(x_1) + b_{\mu i}\eta_i(x_2) = 0 \qquad (\mu = 1, \cdots, p).$$

Let x_3 be a point on x_1x_2 which is not conjugate to either x_1 or x_2. We may suppose that x_3 is the midpoint of x_1x_2 since this can be brought about by a transformation of the form $x = \varphi(\bar{x})$. Suppose that $x_1 = 0$ and let η_i be an arbitrary admissible arc. We now set

$$u_i(x) = \eta_i(x/2), \qquad v_i(x) = \eta_i(x_3 + x/2),$$

$$(w) = (u_1, \cdots, u_n, v_1, \cdots, v_n),$$

$$2q^*(w_1, w_2) = 2q[u(x_1), v(x_2)],$$

$$2\omega^*(x, w, w') = 2\omega(x/2, u, 2u') + 2\omega(x_3 + x/2, v, 2v'),$$

$$J^*(w) = 2q^*(w_1, w_2) + \int_{x_1}^{x_2} 2\omega^*(x, w, w')\, dx,$$

$$\Psi^*_\mu(w) = a_{\mu i}u_i(x_1) + b_{\mu i}v_i(x_2) = 0 \qquad (\mu = 1, \cdots, p),$$

$$\Psi^*_{p+i}(w) = v_i(x_1) - u_i(x_2) = 0 \qquad (i = 1, \cdots, n).$$

It is clear that there is a one-to-one correspondence between the arcs η_i satisfying the conditions $\Psi_\mu = 0$ $(\mu = 1, \cdots, p)$ and the arcs w_j satisfying the conditions $\Psi^*_\nu = 0$ $(\nu = 1, \cdots, p + n)$. Moreover, along corresponding arcs we have $J(\eta) = J^*(w)$. It follows from Theorem 8.6 that the type numbers of J and J^* are equal. In order to show that the orders of degeneracy of J and J^* are the same, we note that w_j is an extremal for J^* if and only if its image η_i is a broken extremal for J having its corner at $x = x_3$. One readily verifies that the transversality conditions for $J^*(w)$ written in terms of the functions u_i, v_i defined above are expressible in the form

$$q_{\eta_{i1}}[u(x_1), v(x_2)] + l_\mu a_{\mu i} - \omega_{\eta_i'}[x_1, u(x_1), 2u'(x_1)] = 0,$$

$$q_{\eta_{i2}}[u(x_1), v(x_2)] + l_\mu b_{\mu i} + \omega_{\eta_i'}[x_2, v(x_2), 2v'(x_2)] = 0,$$

$$\omega_{\eta_i'}[x_3, v(x_1), 2v'(x_1)] - \omega_{\eta_i'}[x_3, u(x_2), 2u'(x_2)] = 0.$$

From these equations it follows at once that an extremal w_j for the J^* satisfies the transversality conditions relative to J^* if and only if its image η_i is an extremal for J satisfying the transversality conditions (7.2). Hence we have the following

LEMMA 9.1. *There exists a functional J^* and a set of end conditions $\Psi^*_\nu = 0$ such that the point x_2 is not conjugate to x_1 relative to J^*, and such that the type number and the order of degeneracy of J^* relative to the conditions $\Psi^*_\nu = 0$ are the same as those of J relative to the conditions $\Psi_\mu = 0$.*

A stronger theorem is

LEMMA 9.2. *The functional J^* described in Lemma 9.1 can be chosen so that the order of concavity and the sum of the orders of the conjugate points of x_1 between x_1 and x_2 of J^* relative to the conditions $\Psi^*_\nu = 0$ are the same as those of J relative to $\Psi_\mu = 0$.*

For by a transformation of the form $x = \varphi(\bar{x})$ we may transform the interval x_1x_2 so that there are no conjugate points on the interval $x_3 \leqq x < x_2$, where x_3 is the midpoint of x_1x_2. If we construct a functional $J^*(w)$ as in the proof of Lemma 9.1, then it is clear that this functional has all the properties described in Lemma 9.2 except, possibly, that its order of concavity r^* might be different from the order of concavity r of J. In order to show that $r = r^*$, let η_{ih} $(h = 1, \cdots, q)$ be a maximal set of extremals for J satisfying the conditions $\Psi_\mu = 0$ and having $J(\eta_h a_h) < 0$ for every set $(a) \neq (0)$ and let w_{jh} be the corresponding extremals for J^*. Clearly we have $J^*(w_h a_h) < 0$ for every set of constants $(a) \neq (0)$. Let $\eta_{i\gamma}$ $(\gamma = q + 1, \cdots, r)$ be a maximal set of linearly independent accessory extremals having $\eta_i(x_1) = \eta_i(x_2) = 0$ and such that no proper linear combination of these extremals satisfies the accessory transversality conditions. Let $w_{j\gamma}$ be the corresponding extremals for J^*. It is clear that $J^*(w_\gamma) = 0$ $(\gamma = q + 1, \cdots, r)$. We may suppose further that the extremals $w_{j\alpha}$ have been chosen so that

$$J^*(w_\alpha, w_\beta) = 0 \qquad (\alpha \neq \beta; \alpha, \beta = 1, \cdots, r). \tag{9.1}$$

Consider now the system

$$J^*(w_\alpha, w) = 0 \qquad (\alpha = 1, \cdots, r-1). \tag{9.2}$$

The extremal w_{jr} satisfies these equations, gives J^* the value zero, but does not satisfy the transversality conditions. It follows from Lemma 7.1 that there exists an extremal w_j satisfying the conditions (9.2) and $\Psi^*_\nu = 0$ and having $J^* < 0$. Hence if we replace w_{jr} by w_j we obtain a set $w_{j\beta}$ $(\beta = 1, \cdots, r)$ satisfying the conditions (9.2) and having $q + 1$ of the numbers $J^*(w_\beta, w_\beta)$ negative. In a similar manner we can replace each of the remaining extremals $w_{j\gamma}$ $(\gamma = q + 1, \cdots, r - 1)$ for which $J^* = 0$ by an extremal for which $J^* < 0$. We obtain thereby a set of r extremals $w_{j\beta}$ satisfying the conditions $\Psi^*_\nu = 0$ and having

$$J^*(w_\alpha, w_\beta) = 0 \ (\alpha \neq \beta), \qquad J^*(w_\beta, w_\beta) < 0 \qquad (\alpha, \beta = 1, \cdots, r).$$

It follows that $r \leqq r^*$.

In order to show that $r^* \leqq r$, let $u_{i\gamma}$ $(\gamma = 1, \cdots, h_1)$ be a maximal set of broken extremals for J having their corners at $x = x_3$, having

$$u_{i\gamma}(x_1) = u_{i\gamma}(x_2) = 0$$

and

$$J(u_\gamma, u_\delta) = 0 \ (\gamma \neq \delta), \qquad J(u_\gamma, u_\gamma) \neq 0 \qquad (\gamma, \delta = 1, \cdots, r).$$

According to Corollary 2 of Theorem 5.1 it is clear that $J(u_\gamma, u_\gamma) > 0$. Let $u_{i\tau}$ $(\tau = h_1 + 1, \cdots, h)$ be a maximal set of linearly independent accessory extremals having $u_{i\tau}(x_1) = u_{i\tau}(x_2) = 0$ and satisfying the accessory transversality conditions. Let $\eta_{i\beta}$ $(\beta = 1, \cdots, m)$ be a maximal set of broken extremals having their corners at $x = x_3$, satisfying the conditions $\Psi_\mu = 0$ and having

$$\begin{aligned} &J(u_\gamma, \eta_\beta) = 0 && (\gamma = 1, \cdots, h), \\ &J(\eta_\alpha, \eta_\beta) = 0 \quad (\alpha \neq \beta), \qquad J(\eta_\beta, \eta_\beta) < 0 && (\alpha, \beta = 1, \cdots, m). \end{aligned} \tag{9.3}$$

It is clear that $m = r^*$. We may suppose that the first q of the arcs $\eta_{i\beta}$ form a maximal set of accessory extremals satisfying these conditions. Then the matrix

$$\| \zeta_{i\sigma}(x_3 - 0) - \zeta_{i\sigma}(x_3 + 0) \| \qquad (\sigma = q + 1, \cdots, r^*). \tag{9.4}$$

will have rank $r^* - q$, where $\zeta_{i\beta}$ denotes the values of $\omega_{\eta'_i}$ along $\eta_{i\beta}$. Otherwise there would exist a set of constants a_σ not all zero, such that

$$[\zeta_{i\sigma}(x_3 - 0) - \zeta_{i\sigma}(x_3 + 0)]a_\sigma = 0 \qquad (\sigma = q + 1, \cdots, r^*).$$

It would follow that the arc $\eta_i = \eta_{i\sigma}a_\sigma$ would be an extremal arc having $J < 0$ and independent of the first q extremal arcs $\eta_{i\alpha}$ $(\alpha = 1, \cdots, q)$, contrary to our choice of these extremals. The matrix (9.4) accordingly has rank $r^* - q$.

We shall now show that $r^* - q \leqq n - h$, where h represents the number of extremals in the set $u_{i\gamma}$ chosen above. To do this we recall that $u_{i\gamma}(x_1) = u_{i\gamma}(x_2) = 0$. Since the arcs $\eta_{i\sigma}$ satisfy the conditions (9.3), it is found by the usual integration by parts that

$$J(u_\gamma, \eta_\sigma) = u_{i\gamma}(x_3)\,[\zeta_{i\sigma}(x_3 - 0) - \zeta_{i\sigma}(x_3 + 0)] = 0$$

$$(\gamma = 1, \cdots, h;\ \sigma = q + 1, \cdots, r).$$

Moreover, the matrix $\| u_{i\gamma}(x_3) \|$ has rank h since the point x_3 is not conjugate to x_1. It follows that the matrix (9.4) can have at most $n - h$ columns if it is to have maximum rank. Hence we have $r^* - q \leqq n - h$. But from the choice of the functions $u_{i\gamma}$ it is clear that $n - h = r - q$ and hence that $r^* \leqq r$, as was to be proved.

We can also establish the further interesting result:

LEMMA 9.3. *There exists a functional J^* and a set of end conditions $\Psi^*_\nu = 0$ such that there are no conjugate points of x_1 on x_1x_2 and such that the type number and orders of degeneracy of J^* relative to the conditions $\Psi^*_\nu = 0$ are equal respectively to those of J relative to $\Psi_\mu = 0$.*

The proof is like that of Lemma 9.1. We select points t_ν $(\nu = 0, 1, \cdots, h)$ on x_1x_2 such that

$$t_0 = x_1 < t_1 < \cdots < t_{h-1} < t_h = x_2$$

and such that there are no conjugate points of $t_{\nu-1}$ on the interval $t_{\nu-1}t_\nu$ $(\nu = 1, \cdots, h)$. We map the interval $t_{\nu-1}t_2$ on x_1x_2 by a transformation $t = \varphi_\nu(x)$ and set

$$u_{i\nu}(x) = \eta_i[\varphi_\nu(x)] \qquad (\nu = 1, \cdots, h)$$

$$(w) = (u_{11}, \cdots, u_{n1}, \cdots, u_{1h}, \cdots, u_{nh}),$$

$$2q^*(w_1, w_2) = 2q[u_1(x_1), u_h(x_2)],$$

$$2\omega^*(x, w, w') = \sum_{\nu=1}^{h} 2\omega(\varphi_\nu, u_\nu, u'_\nu/\varphi'_\nu),$$

$$J^*(w) = 2q^*(w_1, w_2) + \int_{x_1}^{x_2} 2\omega^*(x, w, w')\,dx,$$

$$\Psi^*_\mu(w) = a_{\mu i}u_{i1}(x_1) + b_{\mu i}u_{ih}(x_2) = 0 \qquad (\mu = 1, \cdots, p),$$

$$\Psi^*_{i\nu}(w) = u_{i\nu}(x_1) - u_{i,\nu-1}(x_2) = 0 \qquad (i = 1, \cdots, n;\ \nu = 2, \cdots, h).$$

The proof that the functional J^* so defined has the properties described in Lemma 9.3 is like that of Lemma 9.1. Thus we see that we can not only assume without loss of generality that point x_2 is not conjugate to x_1 but also that there are no conjugate points of x_1 on the interval x_1x_2.

10. The existence of minimal sets. We can now establish the following

THEOREM 10.1. *Let r be the order of concavity of J and s the sum of the orders of the conjugate points of x_1 between x_1 and x_2. If the strengthened condition of Legendre holds on x_1x_2, there exists a set of $m = r + s$ natural isoperimetric conditions*

$$J(\xi_\beta, \eta) = 0 \qquad (\beta = 1, \cdots, m) \tag{10.1}$$

such that the inequality

$$J(\xi_\alpha, \xi_\beta)a_\alpha a_\beta < 0 \qquad (\alpha, \beta = 1, \cdots, m) \tag{10.2}$$

is true for every set of constants $(a) \neq (0)$ and such that $J(\eta) \geqq 0$ for every set of admissible variations η_i satisfying the conditions (10.1), *the equality holding only in case η_i is an accessory extremal satisfying the accessory transversality conditions* (7.2).

For in view of Lemma 9.2 we may assume that the point x_2 is not conjugate to x_1. Let

$$J(\xi_\alpha, \eta) = 0 \qquad (\alpha = 1, \cdots, s)$$

be a proper minimal set for the fixed end point case and let η_{ij} $(j = 1, \cdots, r)$ be a maximal set of accessory extremals satisfying the conditions $\Psi_\mu = 0$ and having

$$J(\eta_h, \eta_j)b_h b_j < 0 \qquad (h, j = 1, \cdots, r) \tag{10.3}$$

for every set of constants $(b) \neq (0)$. Let η_i be an admissible arc satisfying the conditions

$$\begin{aligned} J(\xi_\alpha, \eta) &= 0 \qquad (\alpha = 1, \cdots, s), \\ J(\eta_j, \eta) &= 0 \qquad (j = 1, \cdots, r) \end{aligned} \tag{10.4}$$

and let u_i be the accessory extremal joining the end points of η_i. The arc u_i satisfies the conditions (10.4) since $\xi_{i\alpha}(x_1) = \xi_{i\alpha}(x_2) = 0$ and

$$J(\xi_\alpha, u) = \xi_{i\alpha}\omega_{\eta_i'}(u, u')\Big|_{x_1}^{x_2} = 0,$$

$$J(\eta_j, u) = u_i\omega_{\eta_i'}(\eta_j, \eta_j')\Big|_{x_1}^{x_2} = \eta_i\omega_{\eta_i'}(\eta_j, \eta_j')\Big|_{x_1}^{x_2} = J(\eta_j, \eta) = 0,$$

as one readily verifies. The arc $\bar{\eta}_i = \eta_i - u_i$ has $\bar{\eta}_i(x_1) = \bar{\eta}_i(x_2) = 0$ and satisfies the conditions (10.3). It follows from Theorem 5.1 that $J(\bar{\eta}) \geqq 0$, the equality holding only in case $\bar{\eta}_i$ is identically zero since x_2 is not conjugate to x_1. But

$$J(\bar{\eta}) = J(\eta - u) = J(\eta) - 2J(\eta, u) + J(u) = J(\eta) - J(u)$$

since

$$J(\eta, u) = \eta_i \omega_{\eta_i'}(u, u') \Big|_{x_1}^{x_2} = u_i \omega_{\eta_i'}(u, u') \Big|_{x_1}^{x_2} = J(u) .$$

Hence $J(\eta) \geqq J(u)$, the equality holding in case $\eta_i \equiv u_i$. Moreover, if u_i does not satisfy the accessory transversality conditions, we have $J(u) > 0$, since otherwise the set η_{ij} could not be a maximal set of extremals satisfying the conditions (10.3) and $\Psi_\mu = 0$. It follows that $J(\eta) > 0$ unless η_i is an accessory extremal satisfying the accessory transversality conditions. The set of natural isoperimetric conditions (10.4) accordingly has the properties described in the theorem.

COROLLARY 1. *If the strengthened condition of Legendre holds, the type number of J is equal to the order of concavity of J plus the sum of the orders of the conjugate points of x_1 between x_1 and x_2.*[14]

With the help of Theorem 8.6 one obtains the further result:

COROLLARY 2. *Suppose the strengthened condition of Legendre holds and*

$$J(\xi_\alpha, \eta) = 0 \qquad (\alpha = 1, \cdots, s) \tag{10.5}$$

is a proper minimal set for the fixed end point case. Let η_{ih} $(h = 1, \cdots, r)$ be a set of admissible variations satisfying the conditions (10.5) *and having*

$$J(\eta_h, \eta_j)a_h a_j < 0 \qquad (h, j = 1, \cdots, r)$$

for every set of constants $(a) \neq (0)$. The number of variations in a maximal set of such variations is always the same and is equal to the order of concavity of J.

The following corollary will be useful in §13:

COROLLARY 3. *If the strengthened condition of Legendre holds and there are no points conjugate to x_1 on the interval x_1x_2, the type number and order of degeneracy of J are equal respectively to the negative type number and the nullity of the quadratic form $H(z)$ described in Theorem* 7.1.

In case there are conjugate points of x_1 on the interval x_1x_2, the type number and order of degeneracy of J can also be determined by means of a quadratic form. To prove this choose points t_ν $(\nu = 0, 1, \cdots, h)$ on x_1x_2 as in the proof of Lemma 9.3, and let $\eta_{i\beta}$ $(\beta = 1, \cdots, q)$ be a set of $q = (h + 1)n - p$ linearly independent broken extremals satisfying the conditions $\Psi_\mu = 0$ and having their corners on the hyperplanes $x = t_\nu$ $(\nu = 1, \cdots, h - 1)$. Let

$$H_1(z) = J(\eta_\alpha, \eta_\beta)z_\alpha z_\beta \qquad (\alpha, \beta = 1, \cdots, q). \tag{10.6}$$

Interpreted in terms of the functional J^* described in the proof of Lemma 9.3 it is clear that the quadratic form $H_1(z)$ is one of the type described in Corollary 3 for the new functional J^*. Hence we have the further result:

[14] About a year ago Professor Morse stated to Dr. Hestenes that in an identically normal problem of Bolza the negative type number of his index form is equal to the sum of the orders of the conjugate points of x_1 on x_1x_2 plus the negative type number of the associated Mayer form. This is essentially the equivalent of the result described in this corollary. As far as we know, Professor Morse considers only the non-degenerate case.

COROLLARY 4. *If the strengthened condition of Legendre holds, the type number and the order of degeneracy of J are equal respectively to the negative type number and the nullity of the quadratic form* (10.6).

The quadratic form $H_1(z)$ has been used by Morse to characterize the functional J. He defines the negative type number of $H_1(z)$ to be the index of J. Thus we see that the index of J as defined by Morse is identical with the notion of type number of J here defined.

The order of concavity of J can also be determined by a quadratic form even if the point x_2 is conjugate to x_1. For let x_3 be a value of x such that there are no conjugate points of x_1 on $x_3 \leqq x < x_2$. Let $\eta_{i\alpha}$ $(\alpha = 1, \cdots, 3n - p)$ be a set of $3n - p$ linearly independent broken extremals having their corners at $x = x_3$ and satisfying the conditions $\Psi_\mu = 0$ and put

$$H_2(z) = J(\eta_\alpha, \eta_\beta)z_\alpha z_\beta \qquad (\alpha, \beta = 1, \cdots, 3n - p). \tag{10.7}$$

The following theorem follows at once from the proof of Lemma 9.2.

THEOREM 10.2. *If the strengthened condition of Legendre holds, the order of concavity and order of degeneracy of J are equal respectively to the negative type number and nullity of the quadratic form* (10.7).

11. The order of concavity of a problem with respect to a sub-problem.

Consider now a problem P determined by a functional J and a set of end conditions $\Psi_\mu = 0$ $(\mu = 1, \cdots, p)$. Let P^* be the problem obtained by replacing the end conditions $\Psi_\mu = 0$ by $\Psi^*_\nu = 0$ $(\nu = 1, \cdots, p^*)$. If the functions Ψ_μ are linearly dependent on the functions Ψ^*_ν, the problem P^* will be called a *sub-problem of the problem P*.

Let h be the number of accessory extremals $\eta_{i\alpha}$ $(\alpha = 1, \cdots, h)$ in a maximal set satisfying the conditions $\Psi_\mu = 0$ but not the conditions $\Psi^*_\nu = 0$, and having

$$J(\eta_\alpha, \eta_\beta)a_\alpha a_\beta < 0 \qquad (\alpha, \beta = 1, \cdots, h)$$

for every set of constants $(a) \neq (0)$. Let k be the number of linearly independent extremals $\eta_{i\gamma}$ $(\gamma = h + 1, \cdots, h + k)$ in a maximal set, every proper linear combination of which satisfies the end conditions $\Psi^*_\nu = 0$, the transversality conditions for P^* and the conditions $J(\eta_\alpha, \eta) = 0$ $(\alpha = 1, \cdots, h)$, but not the transversality conditions for the problem P. The number $q = h + k$ will be called the *order of concavity of the problem P with respect to the sub-problem P^**. One readily verifies that $q \leqq p^* - p$. It is clear that the order of concavity r of P as defined in §7 is the order of concavity of P with respect to the sub-problem P^* in which the end conditions are of the form $\eta_i(x_1) = \eta_i(x_2) = 0$.

We can now establish the following result:

THEOREM 11.1. *Let m, m^* be the type numbers and r, r^* be the orders of concavity of P, P^* respectively. Let q be the order of concavity of P with respect to P^*. If the strengthened condition of Legendre holds, then*

$$q = r - r^* = m - m^* \leqq p^* - p\,. \tag{11.1}$$

This result is immediate. For, let U_{ij} $(j = 1, \cdots, k_1)$ be a maximal set of linearly independent accessory extremals having $U_{ij}(x_1) = U_{ij}(x_2) = 0$, every proper linear combination of which satisfies the transversality conditions for sub-problem P^* but not those for P. If h, k are the numbers used in the definition of the order of concavity q of P with respect to P^* and r, r^* are the orders of concavity of P, P^* as defined in §7, then clearly

$$k_1 \leqq k, \qquad h + k_1 = r - r^* \tag{11.2}$$

by virtue of the definitions of r, r^*. It follows that $m^* + q \geqq m$, since by corollary to Theorem 10.1 we have $m = r + s$, $m^* = r^* + s$, where s denotes the sum of the orders of the conjugate points of x_1 between x_1 and x_2. Moreover, $m \geqq m^* + q$. For, let $\eta_{i\gamma}(x)$ $(\gamma = 1, \cdots, q)$ be a set of independent accessory extremals of the type used in the definition of q and consider the system

$$\begin{aligned} J(\xi_\alpha, \eta) &= 0 \qquad (\alpha = 1, \cdots, m^*), \\ J(\eta_\gamma, \eta) &= 0 \qquad (\gamma = 1, \cdots, q), \end{aligned} \tag{11.3}$$

the first m^* of which form a proper minimal set for P^*. We may suppose that the functions $\xi_{i\alpha}$, $\eta_{i\gamma}$ have been chosen so that

$$J(\xi_\alpha, \eta_\gamma) = 0,$$

as one readily verifies. The last k functions of the set give the functional J the value zero since they satisfy the end conditions and the transversality conditions for P^*. But since they do not satisfy the transversality conditions for P, each of these k functions can be replaced by an admissible arc for which $J < 0$ and which satisfies the remaining conditions (11.3), as can be seen by an argument like that given in the proof of Theorem 3.2. It follows that $m = m^* + q$, as was to be proved.

The inequality $0 \leqq m - m^* \leqq p^* - p$ has been given by Morse (p. 92).

Assuming that the strengthened condition of Legendre holds we have

COROLLARY. *Suppose the conditions*

$$J(\xi_\alpha, \eta) = 0 \qquad (\alpha = 1, \cdots, m^*) \tag{11.4}$$

form a proper minimal set for the sub-problem P^* of P and let $\eta_{i\gamma}$ $(\gamma = 1, \cdots, q)$ be a set of q admissible variations for the problem P satisfying the conditions (11.4) *and having*

$$J(\eta_\gamma, \eta_\delta)a_\gamma a_\delta < 0 \qquad (\gamma, \delta = 1, \cdots, q)$$

for every set of constants $(a) \neq (0)$. The number of variations of this type in a maximal set is always the same and is equal to the order of concavity of P with respect to P^*.

12. **Oscillation and comparison theorems.** We assume that the strengthened condition of Legendre holds on x_1x_2. The following general oscillation theorem follows at once from Corollary 1 of Theorem 10.1.

THEOREM 12.1. *Let m be the type number and r the order of concavity of $J(\eta)$. There exist exactly $m - r$ conjugate points of x_1 between x_1 and x_2.*

We have:

COROLLARY. *The number s of conjugate points of x_1 between x_1 and x_2 satisfies the inequality*

$$m - 2n + p \leqq s \leqq m. \tag{12.1}$$

This follows because $r \leqq 2n - p$. The relation (12.1) is the best oscillation theorem of this type given heretofore (Morse, p. 95).

The problem determined by the functional J and the end conditions $\Psi_\mu = 0$ ($\mu = 1, \cdots, p$) will be referred to as the problem P. Let P^* be a problem of the same type determined by a functional J^* and a set of end conditions $\Psi_\nu^* = 0$ ($\nu = 1, \cdots, p^*$). The following result is an immediate consequence of the above theorem.

THEOREM 12.2. *Let m, m^* be the type numbers, r, r^* the orders of concavity, and s, s^* the sums of the orders of the conjugate points of x_1 between x_1 and x_2 of J and J^* respectively. Between these numbers we have the relations*

$$m - m^* = r - r^* + s - s^*,$$

$$| m - m^* | \leqq | s - s^* | + \max [2n - p, 2n - p^*].$$

The last relation follows from the fact that $r \leqq 2n - p$, $r^* \leqq 2n - p^*$.

COROLLARY 1. *If $\omega(\eta, \eta') = \omega^*(\eta, \eta')$, then*

$$m - m^* = r - r^*,$$

$$| m - m^* | \leqq \max [2n - p, 2n - p^*] \leqq 2n.$$

Theorem 12.2 is a special case of the following more general theorem. In this theorem we suppose that P_1, P_1^* are sub-problems of P, P^* respectively and denote the type numbers of the problems P, P^*, P_1, P_1^* by m, m^*, m_1, m_1^* and the number of end conditions for these problems by p, p^*, p_1, p_1^* respectively.

THEOREM 12.3. *If q, q^* are the orders of concavity of P, P^* with respect to the sub-problems P_1, P_1^*, then*

$$m - m^* = q - q^* + m_1 - m_1^*,$$

$$| m - m^* | \leqq | m_1 - m_1^* | + \max [p_1 - p, p_1^* - p^*].$$

This result follows at once from Theorem 11.1.

COROLLARY 1. *If the problems P, P^* have a common sub-problem $P_1 \equiv P_1^*$, then*

$$m - m^* = q - q^*,$$

$$| m - m^* | \leqq \max [p_1 - p, p_1 - p^*] \leqq 2n.$$

COROLLARY 2. *If P_1, P_1^* are sub-problems of the problem $P \equiv P^*$, then*

$$m_1 - m_1^* = q^* - q,$$

$$| m_1 - m_1^* | \leqq \max [p_1 - p, p_1^* - p] \leqq 2n.$$

Consider two problems P, P^* having the same end conditions $\Psi_\mu = 0$ $(\mu = 1, \cdots, p)$ and let

$$J'(\eta) = J(\eta) - J^*(\eta).$$

The functional $J'(\eta)$ and the end conditions $\Psi_\mu = 0$ determine a problem which we shall denote by $P' = P - P^*$.

THEOREM 12.4. *If m' is the type number of the difference problem $P' = P - P^*$, we have the relation*

$$m \leqq m^* + m'.$$

The theorem clearly holds if m' is infinite. Suppose now that m' is finite and that the set

$$J'(\xi_\beta, \eta) = 0 \qquad (\beta = 1, \cdots, m') \tag{12.2}$$

is a proper minimal set of natural isoperimetric conditions for the functional $J' = J - J^*$. It follows that $J(\eta) \geqq J^*(\eta)$ for every admissible arc η_i satisfying these conditions. Let $\eta_{i\gamma}$ $(\gamma = 1, \cdots, k)$ be a maximal set of admissible variations satisfying the conditions (12.2) and having

$$J(\eta_\gamma, \eta_\delta)a_\gamma a_\delta < 0 \qquad (\gamma, \delta = 1, \cdots, h).$$

Clearly $m \leqq h + m'$. Moreover, this relation holds if J is replaced by J^*, since the functions $\eta_{i\gamma}$ satisfy the conditions (12.2). Hence $m - m' \leqq h \leqq m^*$, and the theorem is established.

In case $m' = 0$ we have the stronger

THEOREM 12.5. *Let m, m^* be the type numbers and d, d^* the orders of degeneracy of P, P^*, respectively. If $J(\eta) \geqq J^*(\eta)$ for every admissible arc $(\eta) \not\equiv (0)$, then*

$$m^* \geqq m, \qquad m^* + d^* \geqq m + d.$$

Moreover, if $J(\eta) > J^*(\eta)$ for all such variations, then

$$m^* \geqq m + d.$$

The proof of this theorem is like that of Theorem 6.4 in the fixed end point case.

Consider now two problems P, P^* having the same end conditions and having $\omega(\eta, \eta') \equiv \omega^*(\eta, \eta')$. Let

$$D(\eta_1, \eta_2) = q(\eta_1, \eta_2) - q^*(\eta_1, \eta_2).$$

Let h, k denote respectively the negative type numbers of $D(\eta_1, \eta_2)$, $-D(\eta_1, \eta_2)$ subject to the conditions

$$\Psi_\mu = a_{\mu i}\eta_{i1} + b_{\mu i}\eta_{i2} = 0.$$

We have the following theorem:

THEOREM 12.6. *If the problems P, P^* are related as described above, then*

$$m^* - k \leqq m \leqq m^* + h.$$

By virtue of the corollary to Theorem 8.6 it is clear that in this case the type numbers of the difference problems $P - P^*$ and $P^* - P$ are equal respectively to h and k. Hence by Theorem 12.4 we have $m \leqq m^* + h$, $m^* \leqq m + k$, as was to be proved.

This result has been established by Morse by a different method (Theorem 4.3, p. 93). The general comparison theorem given by Morse on page 94 follows at once from Theorems 12.3 and 12.6. Results of this nature have also been established by Hu[15] by still another method.

13. **Functionals varying continuously with a parameter.** Consider now the functional

$$J(\eta, \sigma) = 2q(\eta_1, \eta_2, \sigma) + \int_{x_1}^{x_2} 2\omega(\eta, \eta', \sigma)\, dx,$$

where $2q$ is a symmetric quadratic form in η_{i1}, η_{i2} whose coefficients are continuous in σ for all values of σ, and 2ω is a quadratic form in η_i, η_i'. We shall assume that the functions

$$P_{ik} = P_{ki}, \qquad Q_{ik}, \qquad R_{ik} = R_{ki}, \qquad \frac{\partial Q_{ik}}{\partial x}, \qquad \frac{\partial R_{ik}}{\partial x},$$

belonging to 2ω, are continuous in x and σ for x on x_1x_2 and for all values of σ. Moreover, the coefficients of η_{i1} and η_{i2} in the end conditions

$$\Psi_\mu(\eta, \sigma) = a_{\mu i}(\sigma)\eta_i(x_1) + b_{\mu i}(\sigma)\eta_i(x_2) = 0$$

are assumed to be continuous in σ for all σ and the matrix $\| a_{\mu i} \quad b_{\mu i} \|$ is supposed to have rank p. We shall assume further that the inequality

$$R_{ik}(x, \sigma)\pi_i\pi_k > 0 \qquad (i, k = 1, \cdots, n)$$

holds on x_1x_2 for all values of σ and for every set $(\pi) \neq (0)$. It follows from Corollary 1 of Theorem 10.1 that the type number $m(\sigma)$ of $J(\eta, \sigma)$ relative to the end conditions $\Psi_\mu = 0$ is finite for all values of σ.

We have the following

THEOREM 13.1. *Let $m(\sigma)$ and $d(\sigma)$ be, respectively, the type number and the order of degeneracy of $J(\eta, \sigma)$. For values of σ sufficiently near σ_0 we have*

[15] Loc cit., pp. 425–443.

$m(\sigma) \geqq m(\sigma_0)$. If $d(\sigma_0) = 0$, then $d(\sigma) = 0$, $m(\sigma) = m(\sigma_0)$. If $d(\sigma) = 0$ $(\sigma \neq \sigma_0)$, then

$$(13.1) \qquad m(\sigma_0) \leqq m(\sigma) \leqq m(\sigma_0) + d(\sigma_0).$$

In order to prove the first statement in the theorem, let $\eta_{i\alpha}(x, \sigma)$ be a set of $m(\sigma_0)$ admissible variations continuous in σ, satisfying the conditions $\Psi_\mu(\eta, \sigma) = 0$ for σ near σ_0, and having

$$J[\eta_\alpha(\sigma), \eta_\beta(\sigma), \sigma]a_\alpha a_\beta < 0 \qquad (\alpha, \beta = 1, \cdots, m(\sigma_0))$$

for $\sigma = \sigma_0$, where $J(\xi, \eta, \sigma)$ is the bilinear functional (7.3) formed for the quadratic functional $J(\eta, \sigma)$. This relation clearly holds for σ sufficiently near σ_0 and hence $m(\sigma) \geqq m(\sigma_0)$ by Theorem 8.6.

In order to prove the inequalities (13.1), we first note that if there are no conjugate points of x' on an interval $x'x''$ relative to $J(\eta, \sigma_0)$, the same is true for σ sufficiently near σ_0. This result follows at once by continuity considerations. It follows that the construction made in the proof of Lemma 9.3 is valid for all σ sufficiently near σ_0 and that we may accordingly assume that there are no conjugate points of x_1 on x_1x_2 relative to $J(\eta, \sigma)$ for these values of σ. Let $\eta_{ih}(x, \sigma)$ $(h = 1, \cdots, 2n - p)$ be a set of $2n - p$ extremals for $J(\eta, \sigma)$ varying continuously with σ and satisfying the conditions $\Psi_\mu(\eta, \sigma) = 0$. Let $H(z, \sigma)$ be the quadratic form

$$H(z, \sigma) = J(\eta_h, \eta_j, \sigma)z_h z_j \quad (h, j = 1, \cdots, 2n - p).$$

This quadratic form is continuous in σ. Moreover according to Corollary 3 of Theorem 10.1 the negative type number and nullity of $H(z, \sigma)$ are equal respectively to type number $m(\sigma)$ and order of degeneracy $d(\sigma)$ of $J(\eta, \sigma)$, since there are no conjugate points of x_1 on x_1x_2 for these values of σ. It follows from the theory of quadratic forms that if $d(\sigma_0) = 0$, then $d(\sigma) = 0$ and $m(\sigma) = m(\sigma_0)$ for values of σ sufficiently near σ_0. Moreover, the same theory tells us that if $d(\sigma) = 0$ $(\sigma \neq \sigma_0)$, as σ increases (or decreases) starting from σ_0, the negative type number $m(\sigma)$ of $H(z, \sigma)$ increases by at most the nullity $d(\sigma_0)$ of $H(z, \sigma_0)$. This proves the relations (13.1), and Theorem 13.1 is established. The arguments used in this paragraph are of the type used by Morse in a similar situation.

A value of σ for which $J(\eta, \sigma)$ is degenerate will be called a *point of degeneracy*. Such a point will be counted a number of times equal to the order of degeneracy of $J(\eta, \sigma)$. It is clear that points of degeneracy are identical with the characteristic roots of the boundary value problem

$$L_i(\eta, \sigma) = \omega_{\eta_i} - (d/dx)\omega_{\eta_i'} = 0,$$

$$T_{i1}(\eta, l, \sigma) = 0, \qquad T_{i2}(\eta, l, \sigma) = 0, \qquad \Psi_\mu(\eta, \sigma) = 0,$$

where the transversality conditions $T_{i1} = 0$, $T_{i2} = 0$ for the functional J are defined as in §7.

THEOREM 13.2. *There are at least* $| m(\sigma_2) - m(\sigma_1) |$ *points of degeneracy (characteristic roots) on the interval* $\sigma_1 \leqq \sigma \leqq \sigma_2$.

This result follows at once from the inequality (13.1) if the number of points of degeneracy on this interval is finite.

THEOREM 13.3. *If for every integer k there exists a set of k constants* $\sigma_1 < \sigma_2 < \cdots < \sigma_k$ *such that* $m(\sigma_j) \neq m(\sigma_{j+1})$ $(j = 1, \cdots, k-1)$, *there exist infinitely many points of degeneracy (characteristic roots). If the function* $m(\sigma)$ *oscillates at each point of degeneracy, this criterion is also a necessary condition for the existence of infinitely many points of degeneracy.*

This result follows at once from Theorem 13.2. Thus we see that if the points of degeneracy are to be finite, the function $m(\sigma)$ can have at most a finite number of points of oscillation.

THEOREM 13.4. *If there exists a sequence* $\{\sigma_k\}$ *such that* $\lim_{k\to\infty} m(\sigma_k) = \infty$, *there exist infinitely many points of degeneracy (characteristic roots). This criterion is also necessary if* $m(\sigma)$ *is monotone and oscillates at each point of degeneracy.*

For clearly in this case $m(\sigma)$ must have infinitely many points of oscillation, since otherwise $m(\sigma_k)$ could not become infinite with k.

The following further criterion is useful:

THEOREM 13.5. *If for every integer k there exist a constant* σ_k *and a set of k admissible arcs* $\eta_{i\alpha}$ $(\alpha = 1, \cdots, k)$ *satisfying the conditions* $\Psi_\mu(\eta, \sigma_k) = 0$ *and having*

$$J(\eta_\alpha, \eta_\beta, \sigma_k)a_\alpha a_\beta < 0 \qquad (\alpha, \beta = 1, \cdots, k) \tag{13.2}$$

for every set of constants $(a) \neq (0)$, *then there exist infinitely many points of degeneracy (characteristic roots). This criterion is also necessary if* $m(\sigma)$ *is monotone and oscillates at each point of degeneracy.*

For by virtue of Theorem 8.6 it is clear that $m(\sigma_k) \geqq k$. Hence

$$\lim_{k\to\infty} m(\sigma_k) = \infty .$$

The theorem now follows from Theorem 13.4.

A more special criterion is

COROLLARY 1. *If there exists a point* x_0, *a positive constant* M, *and a set of constants* $(\pi_1, \cdots, \pi_n)$ *such that*

$$\begin{aligned} &\lim_{\sigma\to\infty} R_{ik}(x_0, \sigma)\pi_i\pi_k < M, \\ &\lim_{\sigma\to\infty} | Q_{ik}(x_0, \sigma)\pi_i\pi_k | < M, \\ &\lim_{\sigma\to\infty} P_{ik}(x_0, \sigma)\pi_i\pi_k = -\infty , \end{aligned} \tag{13.3}$$

there exist infinitely many points of degeneracy (characteristic roots).

For from continuity considerations it is clear that the relations (13.3) hold for all values of x on an interval $x'x''$ containing x_0 in its interior. Let k be

an arbitrary integer and divide the interval $x'x''$ into k equal segments h_α ($\alpha = 1, \cdots, k$). Let $\rho_\alpha(x)$ be a function of class C' which is positive on the interval h_α and identically zero elsewhere. Set $\eta_{i\alpha} = \rho_\alpha(x)\pi_i$. By virtue of the relations (13.3) for any x_0 on $x'x''$ it is clear that there exists a constant σ_k such that the condition (13.2) holds for these variations $\eta_{i\alpha}$. The corollary is thereby established.

The following simple criterion has been given by Morse (p. 99) in a somewhat different form.

COROLLARY 2. *Suppose that the functions R_{ik}, Q_{ik} are independent of σ and $P_{ik} = M_{ik} - \sigma N_{ik}$. If the functions N_{ik} are not all identically zero on x_1x_2, there exist infinitely many points of degeneracy (characteristic roots).*

For clearly in this case there exists a value x_0 and a set of constants (π) such that

$$N_{ik}(x_0)\pi_i\pi_k \neq 0,$$

since otherwise the functions N_{ik} would all be identically zero on x_1x_2, the matrix $\| N_{ik} \|$ being symmetric since $P_{ik} = P_{ki}$ for all σ. For this value x_0 and set of constants (π) the relations (13.3) hold for either σ or $-\sigma$. This proves Corollary 2.

The function $m(\sigma)$ will be called a *proper monotonically increasing function of σ if*

$$m(\sigma) + d(\sigma) \leqq m(\sigma') \qquad (\sigma < \sigma'). \tag{13.4}$$

THEOREM 13.6. *If the function $m(\sigma)$ is a proper monotonically increasing function, the points of degeneracy (characteristic roots) are isolated. Moreover for values of σ sufficiently near a fixed value σ_0*

$$m(\sigma) = m(\sigma_0) \quad (\sigma < \sigma_0), \qquad m(\sigma) = m(\sigma_0) + d(\sigma_0) \quad (\sigma > \sigma_0). \tag{13.5}$$

There are exactly $m(\sigma_2) - m(\sigma_1)$ points of degeneracy on the interval $\sigma_1 \leqq \sigma < \sigma_2$. If there exists a value σ_1 such that $m(\sigma_1) = 0$, there are exactly $m(\sigma)$ points of degeneracy less than σ.

For by virtue of the inequality (13.4) it is clear that there can be at most a finite number of points of degeneracy on any interval $\sigma_1 \leqq \sigma \leqq \sigma_2$, otherwise $m(\sigma)$ would become infinite as σ varies from σ_1 to σ_2. This is not the case. It follows that in this case the points of degeneracy must be isolated. The relations (13.5) follow at once from (13.1) and (13.4). The last two statements of the theorem are immediate consequences of the relations (13.5). Theorem 13.5 is thereby established.

THEOREM 13.7. *If $m(\sigma)$ is a proper monotonically increasing function, the criteria described in Theorems 13.3, 13.4 and 13.5 are necessary and sufficient conditions for the existence of infinitely many points of degeneracy.*

This result is immediate.

The existence of properly monotonically increasing functions $m(\sigma)$ is established by the following

THEOREM 13.8. *Suppose that the end conditions $\Psi_\mu = 0$ are independent of σ and that the inequality*

$$J(\eta, \sigma) \geqq J(\eta, \sigma') \qquad (\sigma < \sigma') \tag{13.6}$$

holds for every set of admissible variations $(\eta) \not\equiv (0)$ satisfying the conditions $\Psi_\mu = 0$. Then the function $m(\sigma)$ is a monotonic increasing function of σ. If the points of degeneracy are isolated, then $m(\sigma)$ is properly monotonic. If the condition (13.6) *holds with the equality sign excluded, then $m(\sigma)$ is properly monotonic.*

This result follows at once from Theorem 12.5. Thus we see that the conclusions described in Theorems 13.6 and 13.7 are true for a functional which satisfies the condition (13.6) with the equality sign excluded, at least if the end conditions are independent of σ. This problem has been fully treated by Morse (ch. IV). He makes the added assumption that $m(\sigma) = 0$ for $-\sigma$ sufficiently large. This assumption is not as restrictive as it seems since if one is interested in the characteristic roots greater than a value σ_0, one can always modify the functional for values $\sigma < \sigma_0$ so that this condition is satisfied.

Consider now the functional $J(\eta)$ and the end conditions $\Psi_\mu = 0$ used in the preceding sections. The functional

$$J(\eta, \sigma) = J(\eta) - \sigma \int_{x_1}^{x_2} \eta_i \eta_i \, dx$$

will be called the *characteristic functional* of J. This functional is clearly of the type described in the last statement of Theorem 13.8. Moreover it is well-known that in this case $m(\sigma) = 0$ for $-\sigma$ sufficiently large, provided that the strengthened condition of Legendre holds for $J(\eta)$, as we shall assume. It follows that the characteristic roots of this functional are not only isolated but infinite in number by virtue of Corollary 2 of Theorem 13.5. We have the following result first proved in the general case by Morse:[16]

THEOREM 13.9. *The type number m of $J(\eta)$ is equal to the number of negative characteristic roots of its characteristic functional $J(\eta, \sigma)$ and its order of degeneracy is equal to the order of $\sigma = 0$ as a characteristic root.*

The Euler equation, the transversality conditions, and the end conditions associated with the characteristic functional $J(\eta, \sigma)$ are the equations

$$\begin{gathered} L_i(\eta) - \sigma\eta_i = 0, \\ T_{i1}(\eta, l) = 0, \qquad T_{i2}(\eta, l) = 0, \qquad \Psi_\mu(\eta) = 0, \end{gathered} \tag{13.7}$$

where $L_i(\eta)$, T_{i1}, T_{i2} are the functions defined by equations (7.1) and (7.2). This system is commonly called the *accessory boundary value problem.* The functions η_i associated with a solution (η_i, σ, l_μ) of this system are called *characteristic solutions* corresponding to the characteristic root σ provided these

[16] Cf. Richardson, *Das Jacobische Kriterium der Variationsrechnung und die Oszillationseigenschaften linearer Differentialgleichungen 2-Ordnung,* Mathematische Annalen, vol. 68 (1910), pp. 279–304.

functions are not all identically zero. It follows that if σ_0 is a characteristic root, then the number $d(\sigma_0)$ of linearly independent characteristic solutions corresponding to σ_0 is equal to the order of degeneracy of $J(\eta, \sigma_0)$. We shall call $d(\sigma_0)$ the *order* of the characteristic root σ_0.

Let $J(\xi, \eta)$ be the bilinear functional belonging to $J(\eta)$ and let $\xi_i(x)$ be a characteristic solution corresponding to a characteristic root. If the arc η_i satisfies the conditions $\Psi_\mu = 0$, then it can be shown that

$$J(\xi, \eta) = \sigma \int_{x_1}^{x_2} \xi_i \eta_i \, dx. \tag{13.8}$$

This result follows at once by the use of equations (13.7). For, let l_μ be the constants with which $\xi_i(x)$, σ satisfy these equations. We then have

$$\begin{aligned} J(\xi, \eta) &= Q(\xi, \eta) + \int_{x_1}^{x_2} \Omega(\xi, \eta) \, dx \\ &= Q(\xi, \eta) + l_\mu \Psi_\mu(\eta) + [\eta_i \omega_{\xi_i'}]_1^2 + \int_{x_1}^{x_2} \eta_i L_i(\xi) \, dx \\ &= T_{i1}(\xi, l)\eta_i(x_1) + T_{i2}(\xi, l)\eta_i(x_2) + \sigma \int_{x_1}^{x_2} \eta_i \xi_i \, dx \\ &= \sigma \int_{x_1}^{x_2} \xi_i \eta_i \, dx, \end{aligned}$$

as was to be proved. Thus we see that if $\sigma \neq 0$, the orthogonality condition

$$\int_{x_1}^{x_2} \xi_i \eta_i \, dx = 0 \tag{13.9}$$

is equivalent to the natural isoperimetric condition $J(\xi, \eta) = 0$. The condition (13.9) has been widely used in the study of boundary value problems.

Let

$$\sigma_1 \leqq \sigma_2 \leqq \cdots \leqq \sigma_m$$

be the negative characteristic roots, each repeated a number of times equal to its multiplicity, and let

$$\xi_{i1}(x), \xi_{i2}(x), \cdots, \xi_{im}(x)$$

be a set of corresponding characteristic solutions. It is understood, of course, that the solutions which correspond to the same characteristic root are linearly independent. It is well known that $J(\eta) \geqq 0$ in the class of admissible arcs $\eta_i(x)$ which satisfy the conditions $\Psi_\mu = 0$ and

$$\int_{x_1}^{x_2} \xi_{i\alpha} \eta_i \, dx = 0 \qquad (\alpha = 1, \cdots, m).$$

We shall give here a new proof of this fact by showing that the conditions

$$J(\xi_\alpha, \eta) = 0 \qquad (\alpha = 1, \cdots, m)$$

form a proper minimal set of natural isoperimetric conditions for $J(\eta)$. In view of equations (13.8) and Theorem 13.9 it is sufficient to show that

$$J(\xi_\alpha, \xi_\beta)a_\alpha a_\beta < 0 \qquad (\alpha, \beta = 1, \cdots, m) \tag{13.10}$$

for every set of constants $(a) \neq (0)$. To prove this we first note that

$$0 = J(\xi_\alpha, \xi_\beta) - J(\xi_\beta, \xi_\alpha) = (\sigma_\alpha - \sigma_\beta)\int_{x_1}^{x_2} \xi_{i\alpha}\xi_{i\beta}\,dx$$

and hence that $J(\xi_\alpha, \xi_\beta) = 0$ whenever $\sigma_\alpha \neq \sigma_\beta$, by virtue of equations (13.8) again. Moreover, as is easily seen, we may suppose that solutions $\xi_{i\alpha}$ have been chosen so that $J(\xi_\alpha, \xi_\beta) = 0$ even in case $\sigma_\alpha = \sigma_\beta$ $(\alpha \neq \beta)$. The inequality (13.10) now follows at once since $\sigma_\alpha < 0$ and

$$J(\xi_\alpha, \xi_\alpha) = \sigma_\alpha \int_{x_1}^{x_2} \xi_{i\alpha}\xi_{i\alpha}\,dx \qquad (\alpha \text{ not summed}),$$

by virtue of equations (13.8).[17]

14. The case of one variable end point. Suppose now that the functional J is of the form

$$J(\eta) = 2q(\eta_1) + \int_{x_1}^{x_2} 2\omega(x, \eta, \eta')\,dx, \tag{14.1}$$

where $2q$ is a symmetric quadratic form in the variables η_{i1}, and that the end conditions are of the form

$$\Psi_\mu = a_{\mu i}\eta_i(x_1) = 0, \qquad \eta_i(x_2) = 0. \tag{14.2}$$

The transversality condition then takes the form

$$T_{i1}(\eta, l) = q_{\eta_{i1}} + l_\mu \Psi_{\mu\eta_{i1}} - \omega_{\eta_i'}(x_1) = 0. \tag{14.3}$$

A value $x_3 = x_1$ will be called a *focal point* of the point x_1 if there exists an accessory extremal $\eta_i(x)$ satisfying the conditions

$$\Psi_\mu = 0, \qquad T_{i1}(\eta, l) = 0, \qquad \eta_i(x_3) = 0$$

with a set of constants l_μ and having $(\eta) \not\equiv (0)$ on x_1x_3. By the *order of a focal point* x_3 will be meant the maximum number of linearly independent accessory extremals satisfying these conditions.

[17] See Hickson, *An application of the Calculus of Variations to boundary value problems*, Trans. Amer. Math. Soc., vol. 31 (1929), pp. 563–79; Reid, *A boundary value problem associated with the Calculus of Variations*, American Journal of Mathematics, vol. 54 (1932), pp. 769–780; Hu, loc. cit. Further references are given in these papers.

The following theorem is well-known. We assume that determinant $|R_{ik}|$ is different from zero.

LEMMA 14.1. *Let η_{ik} ($k = 1, \cdots, n$) be a set of n linearly independent accessory extremals satisfying the conditions*

$$(14.4) \qquad \Psi_\mu = 0, \qquad T_{i1}(\eta, l) = 0$$

with a set of constants $l_{\mu k}$. The system η_{ik} forms a conjugate system. Every accessory extremal η_i satisfying the equations (6.1) *at $x = x_1$, with a set of constants l_μ, is expressible linearly with constant coefficients in the form*

$$\eta_i = \eta_{ik} a_k, \qquad l_\mu = l_{\mu k} a_k.$$

Moreover a value x_3 is a focal point of x_1 if and only if $|\eta_{ik}(x_3)| = 0$.

We have the further well-known

LEMMA 14.2. *Every conjugate system η_{ik} satisfies the conditions* (14.4) *for a suitably chosen functional J and end conditions of the form* (14.1) *and* (14.2).

To prove this, let ζ_{ik} be the values of $\omega_{\eta_i'}$ along the extremals η_{ik}. Let $n - p$ be the rank of the matrix $\|\eta_{ik}(x_1)\|$, and suppose that the functions η_{ik} have been chosen so that $\eta_{i\mu}(x_1) = 0$ ($\mu = 1, \cdots, p$). Set $a_{\mu i} = \zeta_{i\mu}(x_1)$. The conditions $\Psi_\mu = a_{\mu i}\eta_i(x_1) = 0$ will be chosen as our end conditions. In order to determine the quadratic form, we make second choice of our conjugate system η_{ik} so that $\zeta_{ik} = \zeta_{ki}$ at $x = x_1$. We then set $2q = \zeta_{ik}(x_1)\eta_{i1}\eta_{k1}$. One readily verifies that the problem so obtained has the properties described in the theorem.

If one uses the conjugate system η_{ik} described in Lemma 14.1 and makes the obvious changes such as replacing conjugate points by focal points, etc., one obtains the following theorem by precisely the same arguments as those used in the proof of Lemmas 5.1, 5.2 and Theorem 5.1.

THEOREM 14.1. *Let t be the sum of the orders of the focal points of x_1 between x_1 and x_2. If the strengthened condition of Legendre holds, there exists a set of t natural isoperimetric conditions*

$$J(\xi_\beta, \eta) = 0 \qquad (\beta = 1, \cdots, t)$$

such that the inequality $J(\eta) \geqq 0$ is true for every admissible arc η_i satisfying these conditions, the equality holding only in case η_i is an accessory extremal satisfying the transversality conditions (12.3). *Moreover, the admissible arcs $\xi_{i\beta}$ are such that the inequality*

$$J(\xi_\alpha, \xi_\beta)a_\alpha a_\beta < 0 \qquad (\alpha, \beta = 1, \cdots, t)$$

is true for every set $(a) \neq (0)$.

We have the following immediate corollaries. We assume that the strengthened condition of Legendre holds.

COROLLARY 1. *The type number m of J is equal to the sum of the orders of the focal points of x_1 between x_1 and x_2. The order of degeneracy d of J is equal to the order of x_2 as a focal point of x_1.*

In the following corollary and elsewhere we count a focal point (or a conjugate point) a number of times equal to its order.

COROLLARY 2. *The number of focal points of x_1 between x_1 and x_2 is equal to the number of conjugate points of x_1 between x_1 and x_2 plus the order of concavity of J. To obtain the number of focal points of x_1 on $x_1 < x \leqq x_2$, we add to this sum the order of degeneracy of J.*

This result follows from Corollary 1 of Theorem 10.1.

We may call the order of concavity r of J the order of concavity on x_1x_2 of the conjugate system η_{ik} determined by J. Since every conjugate system determines a functional J by virtue of Lemma 14.2, we have the following further results.

COROLLARY 3. *Let t, t^* and r, r^* denote respectively the number of focal points and the order of concavity of two conjugate systems η_{ik}, η^*_{ik} on an open interval x_1x_2. Then*

$$t - t^* = r - r^*.$$

Further results of this nature will be given in the next section.

We shall now give a convenient method of finding the orders of concavity and degeneracy of J. To do so, let η_{ik} be the conjugate system determined by the conditions (12.4) and let u_{ik} be the conjugate system having $u_{ik}(x_2) = 0$. Let ζ_{ik}, v_{ik} be respectively the values of $\omega_{\eta'_i}$ along the extremals η_{ik}, u_{ik}. We have the following preliminary result:

LEMMA 14.3. *If h is the rank of the matrix*

$$\| \zeta_{ij}u_{ik} - \eta_{ij}v_{ik} \|, \tag{14.5}$$

then $n - h$ is equal to the order of degeneracy of J, that is, the order of x_2 as a focal point of x_1.

For since the elements of this matrix (14.5) are all constants, we may evaluate them at $x = x_2$. It is found then that the rank of this matrix is equal to that of the matrix $\| \eta_{ik}(x_2) \|$, since at $x = x_2$ we have $u_{ik} = 0$ and $| v_{ik} | \neq 0$. The lemma now follows from Lemma 14.1 and Corollary 1.

The following theorem gives a method of computing the order of concavity of J. We assume, of course, that the strengthened condition of Legendre holds.

THEOREM 14.2. *Let x_3 be a value such that there are no focal points of x_1 or conjugate points of x_2 on the interval $x_1 < x \leqq x_3$. If $\eta_{ik}(x_3) = u_{ik}(x_3)$, the order of concavity and the order of degeneracy of J are equal respectively to the negative type number and the nullity of the quadratic form*

$$(\zeta_{ij}u_{ik} - \eta_{ij}v_{ik})a_ja_k \qquad (i, j, k = 1, \cdots, n). \tag{14.6}$$

To prove this let

$$J(\xi_\beta, \eta) = 0 \qquad (\beta = 1, \cdots, s) \tag{14.7}$$

be a proper minimal set for the fixed end point case. We may assume that the functions $\xi_{i\beta}$ are identically zero on the interval x_1x_3. Let $\eta_{i\sigma}$ $(\sigma = 1, \cdots, r)$

be a maximal set of admissible variations satisfying the conditions (14.7) and having

$$J(\bar\eta_\sigma, \bar\eta_\tau)b_\sigma b_\tau < 0 \qquad (\sigma, \tau = 1, \cdots, r) \tag{14.8}$$

for every set of constants $(b) \neq (0)$. Suppose the functions η_{ik}, u_{ik} have been chosen so that

$$\bar\eta_{i\sigma}(x_3) = \eta_{i\sigma}(x_3) = u_{i\sigma}(x_3) \qquad (\sigma = 1, \cdots, r).$$

If now we set

$$\bar u_{i\sigma} = \eta_{i\sigma} \quad (x_1 \leqq x \leqq x_3), \qquad \bar u_{i\sigma} = u_{i\sigma} \quad (x_3 \leqq x \leqq x_2),$$

it is clear that the inequality (14.8) is still true if we replace the functions $\bar\eta_{i\sigma}$ by $\bar u_{i\sigma}$ since the set (14.7) forms a minimal set for the fixed end point case. Moreover, by the usual integration by parts and the use of the transversality conditions it is found that along the arc $\eta_i = \bar u_{i\sigma} b_\sigma$ we have

$$\begin{aligned} J(\eta) &= 2q + \int_{x_1}^{x_3} 2\omega \, dx + \int_{x_3}^{x_2} 2\omega \, dx \\ &= \eta_i(x_3 - 0)\zeta_i(x_3 - 0) - \eta_i(x_3 + 0)\zeta_i(x_3 + 0) \\ &= \eta_i(x_3 + 0)\zeta_i(x_3 - 0) - \eta_i(x_3 - 0)\zeta_i(x_3 + 0) \\ &= (\zeta_{i\sigma} u_{i\tau} - \eta_{i\sigma} v_{i\tau})b_\sigma b_\tau \qquad (\sigma, \tau = 1, \cdots, r). \end{aligned}$$

One readily verifies that every broken extremal satisfying the end conditions (14.2) and having its corners at $x = x_3$ also satisfies the conditions (14.7). It follows that the negative type number of the quadratic form (14.6) is equal to r. The nullity of this form is clearly equal to $n - h$, where h is the rank of matrix (14.5), and the theorem is proved.

15. The case of two end manifolds. Consider now the functional

$$J = 2q_1(\eta_1) - 2q_2(\eta_2) + \int_{x_1}^{x_2} 2\omega(\eta, \eta') \, dx, \tag{15.1}$$

where $2q_1$ and $2q_2$ are symmetric quadratic forms in the variables η_{i1} and η_{i2}, respectively. The end conditions are assumed to be of the form

$$\begin{aligned} \Psi_{\mu 1}(\eta_1) &= a_{\mu i}\eta_i(x_1) = 0 \qquad (\mu = 1, \cdots, p_1 \leqq n), \\ \Psi_{\nu 2}(\eta_2) &= b_{\nu i}\eta_i(x_2) = 0 \qquad (\nu = 1, \cdots, p_2 \leqq n). \end{aligned} \tag{15.2}$$

In this case the transversality conditions are expressible in the form

$$T_{i1}(\eta_1, l_1) = q_{1\eta_{i1}} + l_{\mu 1}a_{\mu i} - \omega_{\eta_i'}(x_1) = 0,$$

$$T_{i2}(\eta_2, l_2) = q_{2\eta_{i2}} + l_{\nu 2}b_{\nu i} - \omega_{\eta_i'}(x_2) = 0,$$

as one readily verifies.

The problem determined by the system (15.1) and (15.2) will be called the problem P. The problems obtained by replacing the end conditions $\Psi_{\nu 2} = 0$ by $\eta_i(x_2) = 0$ will be called the problem P_1. By the problem P_2 will be meant the problem obtained by replacing the end conditions $\Psi_{\mu 1} = 0$ by the conditions $\eta_i(x_1) = 0$. The problems P_1 and P_2 are of the type described in the last section. By the use of Theorem 11.1 we obtain the following important

THEOREM 15.1. *Let m, m_1, m_2 be respectively the type numbers of the problems P, P_1, P_2, and let q_1, q_2 be the orders of concavity of P with respect to P_1 and P_2. If the strengthened condition of Legendre holds, then*

$$m = m_1 + q_1 = m_2 + q_2. \tag{15.3}$$

The transversality conditions and the end conditions for the problem P_1 are given by the system

$$\Psi_{\mu 1}(\eta_1) = 0, \qquad T_{i1}(\eta_1, l_1) = 0, \tag{15.4}$$

and for the problem P_2 by the system

$$\Psi_{\nu 2}(\eta_2) = 0, \qquad T_{i2}(\eta_2, l_2) = 0. \tag{15.5}$$

Let η^1_{ik}, η^2_{ik} be the conjugate systems of accessory extremals determined by (15.4), (15.5) respectively. It will be convenient to denote these systems by H_1 and H_2. According to Lemma 14.2 every pair of conjugate systems H_1, H_2 determine a pair of problems P_1, P_2, which are obviously sub-problems of a problem P defined by a functional of the type (15.1) subject to end conditions of the form (15.2). It follows that the general theory concerning pairs of conjugate systems is obtained by the study of the problem P.

Let ζ_i be the value of $\omega_{\eta'_i}$ along the arc η_i. It is well-known that if an extremal satisfies the conditions

$$\zeta_{ij}\eta_i - \eta_{ij}\zeta_i = 0,$$

where η_{ij} is a conjugate system, then η_i is linearly dependent with constant coefficients on η_{ij}. Hence we have the following result.

LEMMA 15.1. *If h is the rank of the matrix*

$$\| \zeta^1_{ij}\eta^2_{ik} - \eta^1_{ij}\zeta^2_{ik} \|,$$

the order of degeneracy of J is equal to $n - h$.

Thus we see that the order of degeneracy of J is equal to the maximum number of linearly independent extremals which the systems H_1, H_2 have in common.

The zero x_3 of the determinant $| \eta_{ik}(x) |$ will be called a focal point of the conjugate system η_{ik} and its order will be defined to be $n - h$, where h is the rank of the matrix $\| \eta_{ik}(x_3) \|$. A focal point will be counted a number of times equal to its order.

Suppose now that we have given two conjugate systems H_1, H_2, as above.

Let d denote the maximum number of linearly independent extremals which these systems have in common. Let x_3 be a value of x which is not a focal point of either H_1 or H_2 and choose the functions η^1_{ik}, η^2_{ik} belonging to H_1, H_2 respectively so that $\eta^1_{ik} = \eta^2_{ik}$ at $x = x_3$. Let $q(x_3)$ be the negative type number of the quadratic form

$$(\zeta^1_{ij}\eta^2_{ik} - \eta^1_{ij}\zeta^2_{ik})a_j a_k \qquad (i, j, k = 1, \cdots, n). \tag{15.6}$$

The nullity of this quadratic form is clearly equal to d. The function $q(x)$ is well defined except at focal point x_3 of H_1 or H_2. At such a point we define $q(x_3 - 0)$, $q(x_3 + 0)$ to be the left hand and right hand limits of $q(x)$. It is clear that $q(x)$ is a constant on any interval which contains no focal point of either H_1 or H_2. The function $q(x)$ will be called the *index function* of H_1 with respect to H_2. The index function of H_2 with respect to H_1 is clearly the function $n - d - q(x)$. It follows that $q(x) \leqq n - d$.

Suppose that H_1, H_2 belong to the problems P_1, P_2, P described in the paragraph preceding Theorem 15.1. We have the following result.

LEMMA 15.2. *The number $q(x_1 + 0)$ is equal to the order of concavity of the problem P with respect to the sub-problem P_2. The number $q(x_2 - 0)$ is equal to the order of concavity of P with respect to P_1.*

The proof of the first statement of the theorem is precisely that of Theorem 14.2 except for obvious changes in notation. In this case we choose the set (14.7) to be a proper minimal set for the problem P_2 such that the functions are identically zero on an interval $x_1 < x \leqq x_3$ which contains no focal points of x_1 or x_2. If we replace the conjugate system η_{ik}, u_{ik} by η^1_{ik}, η^2_{ik}, it is found that $q(x_3) = q(x_1 + 0)$ is equal to the order of concavity of P with respect to P_2. The last statement follows similarly.

The following important result can now be proved.

THEOREM 15.2. *Let H_1, H_2 be two conjugate systems and $q(x)$ be the index function of H_1 with respect to H_2. Let t_1, t_2 denote the number of focal points of H_1, H_2, respectively, on an interval γ. Then*

$$\begin{aligned} t_1 - t_2 &= q(x_1 + 0) - q(x_2 - 0) \quad \text{on} \quad (x_1 < x < x_2), \\ t_1 - t_2 &= q(x_1 + 0) - q(x_2 + 0) \quad \text{on} \quad (x_1 < x \leqq x_2), \\ t_1 - t_2 &= q(x_1 - 0) - q(x_2 - 0) \quad \text{on} \quad (x_1 \leqq x < x_2), \\ t_1 - t_2 &= q(x_1 - 0) - q(x_2 + 0) \quad \text{on} \quad (x_1 \leqq x \leqq x_2). \end{aligned} \tag{15.7}$$

In each of these cases $| t_1 - t_2 | \leqq n - d$, where d is the maximum number of linearly independent extremals belonging to both H_1 and H_2.

For, the systems H_1, H_2 determine problems P, P_1, P_2 of the type described above. The first of the relations (15.6) follows at once from Lemma 15.2 and the relations (15.3) since $m_1 = t_1$, $m_2 = t_2$ in this case. The remaining relations follow at once from the first by increasing x_2, or decreasing x_1, slightly. The last statement is a consequence of the fact that $0 \leqq q(x) \leqq n - d$, as was seen above.

Thus we see that the equations (15.6) give us a method of finding the number of focal points of a conjugate system H_1 on an interval γ when the number of focal points of a second system H_2 are known. In particular, if the number of conjugate points of x_1 on x_1x_2 is known, the number of focal points of any conjugate system η_{ik} can be computed by the use of equations (15.7).

The quadratic form (15.6) has appeared in the literature in many different forms. It has been used effectively in the two dimensional case by Bliss,[18] and in the n-dimensional case by Morse,[19] Currier,[20] and Hestenes.[21] This quadratic form has been used by Morse to obtain the inequality $| t_1 - t_2 | \leqq n - d$ and other stronger inequalities.[22] The relations (15.7) were first given by Currier.[20]

16. **Periodic extremals.** Suppose now that the strengthened condition of Legendre holds and that there exists a constant τ and a set of n^2 constants A_{ij}, with $A = | A_{ij} | \neq 0$, such that the equation

$$2\omega(x, \eta, \eta') = 2\omega(x + \tau, u, u') \tag{16.1}$$

is an identity in x, η_i, η_i', where

$$u_i = A_{ij}\eta_j, \qquad u_i' = A_{ij}\eta_j'.$$

It follows that the expression

$$L_i(x, \eta, \eta', \eta'') = A_{ji}L_j(x + \tau, u, u', u''), \tag{16.2}$$

where

$$L_i = \omega_{\eta_i} - (d/dt)\omega_{\eta_i'}, \qquad u_i'' = A_{ij}\eta_j'',$$

is an identity in x, η_i, η_i', η_i''. The identities (16.1) and (16.2) are still valid if we replace τ by $\alpha\tau$ and A_{ij} by A_{ij}^α, where α is an integer and A_{ij}^α are the elements of the matrix A^α.

LEMMA 16.1. *If η_{ik} is a conjugate system of accessory extremals and*

$$u_{ik}(x + \alpha\tau) = A_{ij}^\alpha \eta_{jk}(x),$$

the functions $u_{ik}(x)$ define a conjugate system. Moreover, the number of focal points of the conjugate system u_{ik} on the interval $x_1 + \alpha\tau < x \leqq x_2 + \alpha\tau$ is equal to the number of focal points of η_{ik} on $x_1 < x \leqq x_2$.

[18] Bliss, *A boundary value problem of the Calculus of Variations*, Bulletin of the American Mathematical Society, vol. 32 (1926), pp. 317–331.

[19] Morse, pp. 65–70, 106. See also *A generalization of the Sturm separation and comparison theorems in n-space*, Mathematische Annalen, vol. 103 (1930), pp. 52–69.

[20] Currier, *The variable end point problem of the Calculus of Variations including a generalization of the classical Jacobi condition*, Transactions of the American Mathematical Society, vol. 34 (1932), pp. 689–704.

[21] Hestenes, *Sufficient conditions for the problem of Bolza in the Calculus of Variations*, Transactions of the American Mathematical Society, vol. 36 (1934), pp. 793–818.

[22] In this connection one should also see Hu, loc. cit.

The functions u_{ik} $(k = 1, \cdots, n)$ are clearly extremals by virtue of the identity (16.2) with τ replaced by $\alpha\tau$ and A_{ij} by A^{α}_{ij}. If a similar replacement is made in equations (16.1), by differentiating with respect to η'_i it is found that

$$\omega_{\eta'_i}(x, \eta, \eta') = A^{\alpha}_{ji}\omega_{\eta'_j}(x + \alpha\tau, u, u').$$

If we denote the values of $\omega_{\eta'_i}$ along η_{ik}, u_{ik} by ζ_{ik}, v_{ik}, it is clear from this last expression that

$$\eta_{ij}(x)\zeta_{ik}(x) = u_{ij}(x + \alpha\tau)v_{ik}(x + \alpha\tau),$$

and hence that

$$0 = \eta_{ij}\zeta_{ik} - \eta_{ik}\zeta_{ij} = u_{ij}v_{ik} - u_{ik}v_{ij}.$$

The extremals u_{ik} accordingly form a conjugate system. The last statement in the lemma is immediate.

An extremal η_i will be said to be periodic if there exists an integer α such that

$$\eta_i(x + \alpha\tau) = A^{\alpha}_{ij}\eta_j(x), \qquad \zeta_i(x) = A^{\alpha}_{ji}\zeta_j(x + \alpha\tau). \tag{16.3}$$

If α is the smallest integer for which this condition holds, then η_i will be said to be of period $\alpha\tau$. It is clear that if the relation (16.3) hold for one value of x it holds for all values of x by virtue of the identities (16.1) and (16.2).

Let η_{ik} be a conjugate system and $t_h(x_1)$ be the number of focal points of η_{ik} on the interval $x_1 < x \leqq x_1 + h\tau$. Let d_h be the maximum number of periodic extremals which are linearly dependent on the system η_{ik} and whose periods are divisors of $h\tau$. We have the following[23]

THEOREM 16.1. *The limit*

$$\mu = \lim_{h \to \infty} \frac{t_h(x_1)}{h} \tag{16.4}$$

exists, is independent of x_1, and is the same for all conjugate systems η_{ik}. Moreover, the inequalities

$$| h\mu - t_h(x_1) | \leqq n - d_h \qquad (h = 1, 2, \cdots) \tag{16.5}$$

are true for all values of x_1. If $\eta_{ik}(x_1) = 0$, then

$$t_h(x_1) \leqq h\mu \leqq t_h(x_1) + n - d_h \qquad (h = 1, 2, \cdots). \tag{16.6}$$

For, let $\alpha = hk$ and

$$u_{ik}(x + \alpha\tau) = A^{\alpha}_{ij}\eta_{jk}(x).$$

[23] Cf. Morse and Pitcher, *On certain invariants of closed extremals*, Proceedings of the National Academy of Sciences, vol. 20 (1934), pp. 282–287. See also Hedlund, *Poincaré's rotation number and Morse's type number*, Transactions of the American Mathematical Society, vol. 34 (1932), pp. 75–97.

The number of focal points of $u_{ik}(x)$ on $x_1 + (\alpha - h)\tau < x \leqq x_1 + \alpha\tau$ is equal to $t_h(x_1)$, by Lemma 16.1. According to Theorem 15.2 the number of focal points of η_{ik} on this interval differs from $t_h(x_1)$ by at most $n - d_h$. This statement holds for all values of the integer k in $\alpha = hk$ and hence

$$| t_\alpha(x_1) - kt_h(x_1) | \leqq k(n - d_h),$$
$$| t_\alpha(x_1) - ht_k(x_1) | \leqq h(n - d_k).$$

If now we divide by $\alpha = hk$, we find that

$$\text{(16.7)}\qquad \begin{aligned} 0 &\leqq \left| \frac{t_\alpha(x_1)}{\alpha} - \frac{t_h(x_1)}{h} \right| \leqq (n - d_h)/h \leqq n/h, \\ 0 &\leqq \left| \frac{t_\alpha(x_1)}{\alpha} - \frac{t_k(x_1)}{k} \right| \leqq (n - d_k)/k \leqq n/k, \end{aligned}$$

and hence that

$$\left| \frac{t_h(x_1)}{h} - \frac{t_k(x_1)}{k} \right| \leqq n/h + n/k.$$

These relations hold for all values of x_1. For h, k sufficiently large, the right member in the last expression can be made as small as we please. It follows that the limit μ in (16.4) exists and is the same for all values of x_1. Since the numbers $t_h(x_1)$ for two different conjugate systems differ by at most n, it is clear that the constant μ is the same for all conjugate systems η_{ik}. The relation (16.5) follows from the first of the relations (16.7) by letting the integer k in $\alpha = hk$ become infinite. If $\eta_{ij}(x_1) = 0$, the expressions (16.7) are still valid if the absolute value signs are removed. The relation (16.6) is then obtained by letting k become infinite again. This proves Theorem 16.1.

The constant μ will be called the frequency number:

COROLLARY. *If there exists a conjugate system η_{ik} all of whose extremals are periodic, then the frequency number μ is rational.*[24]

For, in this case the constant d_h appearing in (16.5) is equal to n for a suitably chosen integer h. For this value of h we have $h\mu = t_h(x_1)$. It follows that μ must be rational.

[24] By the use of a theorem of Birkhoff on linear differential equations with periodic coefficients (*Dynamical Systems*, pp. 77–89) it can be shown immediately that all conjugate points of x_1 satisfy an equation $\Delta(x, x_1) = 0$, where in the general case $\Delta(x_1, x) = \Delta(x, x_1)$ is a linear homogeneous expression of the form

$$\Sigma\, p(x)\, e^{(\pm \lambda_1 \pm \lambda_2 + \cdots \pm \lambda_n)x},$$

$p(x)$ is periodic in x of period τ and $\lambda_1, \cdots, \lambda_n$ are n quantities, real or complex. By means of this form complete results on the asymptotic distribution of the conjugate points can be obtained. Results of this sort have been announced by Morse and Pitcher (loc. cit.) without indication of the method employed.

III

The general fixed end point problem in non-parametric form

In §§17–21 we shall be concerned with the characterization of extremals of an integral of the form

$$J = \int_{x_1}^{x_2} f(x, y_1, \cdots, y_n, y_1', \cdots, y_n')\, dx = \int_{x_1}^{x_2} f(x, y, y')\, dx$$

in $(x_1, y_1, \cdots, y_n)$-space. Among other things, it will be shown that if there are m conjugate points of the initial point 1 on an extremal arc E_{12}, there exists a set of m natural isoperimetric conditions (and no fewer) such that E_{12} affords a minimum to J relative to neighboring admissible arcs joining the end points of E_{12} and satisfying these m natural isoperimetric conditions, provided, of course, that the strengthened conditions of Weierstrass and Legendre hold on E_{12} and the point 2 is not conjugate to 1. The strengthened condition of Weierstrass here used is a modification of the classical one and reduces to the latter in case there are no conjugate points of the point 1 on E_{12}. This result is established without the use of the classical isoperimetric theory. This is possible only because of the special properties of natural isoperimetric conditions.

17. Hypotheses and preliminary results. It is assumed that the function $f(x, y, y')$ is of class C^4 in an open region $\mathfrak{R}$ of points (x, y, y'). A set of elements (x, y, y') will be called admissible if it is in $\mathfrak{R}$. By an *admissible arc* will be meant an arc

$$y_i = y_i(x) \qquad (x_1 \leqq x \leqq x_2,\ i = 1, \cdots, n)$$

of class D', all of whose elements (x, y, y') are in $\mathfrak{R}$, and joining two fixed points (x_1, y_{i1}) and (x_2, y_{i2}).

An arc of class C'' with elements (x, y, y') in $\mathfrak{R}$ which satisfies the *Euler equations*

$$(17.1) \qquad P_i = f_{y_i} - (d/dx)f_{y_i'} = 0 \qquad (i = 1, \cdots, n)$$

will be called an *extremal arc*. Such an arc will be said to be *non-singular* if the determinant

$$R = | f_{y_i' y_k'} |$$

is different from zero at each element (x, y, y') on it. It is well-known that a non-singular extremal arc E_{12} is a member of a $2n$-parameter family of extremal arcs

$$(17.2) \qquad y_i = y_i(x, c_1, \cdots, c_{2n})$$

for special values $x_1 \leqq x \leqq x_2$, $c_s = c_{s0}$. The functions $y_i(x, c)$, $y_{ix}(x, c)$ are of class C^3 in a neighborhood of the values (x, c) on E_{12} and the determinant

$$\begin{vmatrix} y_{ic_s} \\ y_{ixc_s} \end{vmatrix} \qquad (s = 1, \cdots, 2n)$$

is different from zero on E_{12}.

Along an extremal arc E_{12} the second variation of the integral J is expressible in the form

$$J_2(\eta) = \int_{x_1}^{x_2} 2\omega(x, \eta, \eta')\, dx,$$

where

$$2\omega = f_{y_i y_k} \eta_i \eta_k + 2f_{y_i y_k'} \eta_i \eta_k' + f_{y_i' y_k'} \eta_i' \eta_k' .$$

The variations $\eta_i(x)$ which we admit are of class D' and vanish at x_1 and x_2. Thus we see that the study of the second variations leads us to a problem of the type described in §§1–6.

A point 3 on E_{12} will be said to be conjugate to the point 1 on E_{12} if it is defined by a value x_3 conjugate to x_1 in the sense described in §1. By the order of the point 3 as a conjugate point of 1 will be meant the order of x_3 as a conjugate point of x_1. It is well-known that the conjugate points of x_1 are defined by the zeros $x_3 \neq x_1$ of the determinant

$$\begin{vmatrix} y_{ic_s}(x, c_0) \\ y_{ic_s}(x_1, c_0) \end{vmatrix}$$

belonging to the family (17.2). This result follows because the functions $\eta_{is} = y_{ic_s}(x, c)$ form a set of $2n$ linearly independent accessory extremals for the functional $J_2(\eta)$. An important consequence of this fact is that if the point 2 on E_{12} is not conjugate to the point 1, then every pair of points which lie respectively in a sufficiently small neighborhood of the end points of E_{12} can be joined by a unique extremal arc. The conjugate points of the point 1 are also determined by the zero $x_3 \neq x_1$ of the determinant

$$| y_{ib_k}(x, b_0) |$$

belonging to an n-parameter family of extremals

$$y_i = y_i(x, b_1, \cdots, b_n)$$

passing through the point 1 on E_{12}, containing E_{12} for the values $x_1 \leqq x \leqq x_2$, $b_i = b_{i0}$, and having its determinant $| y_{ixb_k} |$ different from zero at the point 1 on E_{12}. The continuity properties of this family are, of course, the same as those of the family (17.2). This result follows from the remarks made in §1 since in this case the functions $\eta_{ik}(x) = y_{ib_k}(x, b)$ form a set of n linearly independent accessory extremals having $\eta_{ik}(x_1) = 0$.

18. **Natural isoperimetric conditions.** Let $\xi_i(x)$ be a function of class C^3 on x_1x_2 and vanishing at x_1 and x_2. The condition

$$J_1(\xi) = \int_{x_1}^{x_2} \{f_{y_i}\xi_i + f_{y'_i}\xi'_i\}\, dx = 0$$

will be called a *natural isoperimetric condition.* This condition is satisfied by every arc which satisfies the equations

$$f_{y'_i} = \int_{x_1}^{x} f_{y_i}\, dx + c_i \qquad (i = 1, \cdots, n) \tag{18.1}$$

with a set of constants c_i, and hence also by every extremal arc. This fact is readily established by integration by parts. Conversely, an admissible arc which satisfies all natural isoperimetric conditions is necessarily a solution of equations (18.1). This result follows from the fundamental lemma in the Calculus of Variations, since in this case $J_1(\eta) = 0$ for every admissible arc η_i of class C^3 and hence also for those of class D'.

It is clear that the function $\xi_i(x)$ used in the definition of natural isoperimetric conditions can be replaced by functions $\xi_i(x, y)$ which vanish at the points 1 and 2. However, for the present purposes, we shall restrict ourselves to functions of the first type.

Consider now a set of m natural isoperimetric conditions

$$J_1(\xi_\alpha) = \int_{x_1}^{x_2} f_\alpha\, dx = 0 \qquad (\alpha = 1, \cdots, m), \tag{18.2}$$

where

$$f_\alpha = \xi_{i\alpha} f_{y_i} + \xi'_{i\alpha} f_{y'_i}.$$

If we set

$$F(x, y, y', \lambda) = f + \lambda_\alpha f_\alpha,$$

the Euler-Lagrange equations of the integral J relative to the conditions (18.2) take the interesting form

$$F_{y_i} - (d/dx)F_{y'_i} = P_i + \lambda_\alpha L_i(\xi_\alpha) = 0, \tag{18.3}$$

where P_i are the functions (17.1) and

$$L_i(\eta) = \omega_{\eta_i} - (d/dx)\omega_{\eta'_i}.$$

An arc

$$y_i = y_i(x) \qquad (x_1 \leqq x \leqq x_2),$$

which satisfies these equations with a set of constant multipliers λ_α will be called an *isoperimetric extremal.* Along such an arc we have by the usual integration by parts

$$J_1(\xi_\alpha) = \int_{x_1}^{x_2} \xi_{i\alpha} P_i\, dx = -\int_{x_1}^{x_2} \xi_{i\alpha} L_i(\xi_\beta)\lambda_\beta\, dx = -J_2(\xi_\alpha, \xi_\beta)\lambda_\beta, \tag{18.4}$$

where $J_2(\xi, \eta)$ is the bilinear functional associated with the quadratic functional $J_2(\eta)$, as described in §2. It follows that if the determinant

$$| J_2(\xi_\alpha, \xi_\beta) | \tag{18.5}$$

is different from zero along an isoperimetric extremal E_{12}, this arc can satisfy the conditions (18.2) only in case its multipliers λ_α are all zero, that is, only in case E_{12} is an extremal of the functional J.

It should be noted that if the determinant

$$| F_{y_i' y_k'} |$$

is different from zero along an isoperimetric extremal E_{12}, then E_{12} is a member of a $(2n + m)$-parameter family of isoperimetric extremals

$$y_i = y_i(x, c_1, \cdots, c_{2n}, \lambda_1, \cdots, \lambda_m) \tag{18.6}$$

with multipliers $\lambda_1, \cdots, \lambda_m$. The functions $y_i(x, c, \lambda)$, $y_{ix}(x, c, \lambda)$ are of class C'' in a neighborhood of the values (x, c, λ) belonging to E_{12} and the determinant

$$\begin{vmatrix} y_{ic_s} \\ y_{ixc_s} \end{vmatrix}$$

is different from zero along E_{12}. It is clear that for fixed values of (λ) the family (18.6) defines the usual extremal family belonging to the integral

$$J_\lambda = \int_{x_1}^{x_2} F\, dx.$$

We shall make use of this fact in our sufficiency proofs. If the isoperimetric extremal E_{12} satisfies the conditions (18.2) and is such that the determinant (18.5) is different from zero on it, then E_{12} is a member of a $2n$-parameter family of isoperimetric extremals satisfying the conditions (18.2) and this family is precisely the extremal family (17.2) for the functional J, as one readily verifies.

19. **Minimal sets.** A set of m natural isoperimetric conditions

$$J_1(\xi_\alpha) = \int_{x_1}^{x_2} f_\alpha\, dx = 0 \qquad (\alpha = 1, \cdots, m) \tag{19.1}$$

will be said to form a *minimal set* for an admissible arc E_{12} satisfying these conditions if the arc E_{12} affords at least a weak minimum to the integral J relative to neighboring admissible arcs satisfying the conditions (19.1), and if no proper subset of these conditions has this property. Such a set will be called a *proper set* if the determinant

$$| J_2(\xi_\alpha, \xi_\beta) | \tag{19.2}$$

is different from zero on E_{12}. In the following pages we shall restrict ourselves to proper minimal sets unless otherwise expressly stated. In this case there

are infinitely many admissible arcs joining the end points of E_{12} and satisfying the conditions (19.1). This result follows from the proof of the following lemma:

LEMMA 19.1. *If the set* (19.1) *forms a proper set for an admissible arc* E_{12} *and* $\eta_i(x)$ *is a set of admissible variations for* E_{12} *satisfying the conditions*

$$J_2(\xi_\alpha, \eta) = 0 \qquad (\alpha = 1, \cdots, m) \tag{19.3}$$

on E_{12}, *there exists a one-parameter family of admissible arcs*

$$y_i = y_i(x, b) \qquad (x_1 \leqq x \leqq x_2) \tag{19.4}$$

satisfying the conditions (19.1), *containing* E_{12} *for* $b = 0$, *and having* $\eta_i = y_{ib}(x, 0)$. *The further derivatives* y_{ixb}, y_{ibb}, y_{ixbb} *exist and are continuous except possibly at a finite number of values of* x *on* x_1x_2.

To prove this, let $\eta_i(x)$ be an arbitrary admissible variation which satisfies the conditions (19.3) along E_{12} and let

$$Y_i(x, a_1, \cdots, a_m, b) = y_i(x) + a_\alpha\xi_\alpha(x) + b\eta_i(x),$$

the functions $y_i(x)$ being those belonging to E_{12}. When these functions are substituted for y_i in the equations (19.1), a set of m functions $J_{1\alpha}(a, b)$ having $J_{1\alpha}(0, 0) = 0$ is obtained. The determinant of the derivatives of $J_{1\alpha}$ with respect to $a_1, \cdots, a_m$ at $(a, b) = (0, 0)$ is equal to the determinant (19.2) and is accordingly different from zero on E_{12}. The equations $J_{1\alpha}(a, b) = 0$ therefore have a unique solution $a_i = A_i(b)$ of class C^3 near $b = 0$ and having $A_\alpha(0) = 0$. Moreover $A'_\alpha(0) = 0$. This is also a consequence of the non-vanishing of the determinant (19.2) since in this case we have

$$\begin{aligned}(d/db)\, J_{1\alpha}[A(b), b]\,\big|_{b=0} &= J_2(\xi_\alpha, \xi_\beta)\, A'_\beta(0) + J_2(\xi_\alpha, \eta) \\ &= J_2(\xi_\alpha, \xi_\beta)\, A'_\beta(0) = 0,\end{aligned}$$

as one readily verifies. It is clear, therefore, that the one-parameter family

$$y_i = y_i(x, b) = Y_i[x, A(b), b]$$

satisfies equations (19.1), contains E_{12} for $b = 0$, and has $\eta_i(x) = y_{ib}(x, 0)$. This proves Lemma 19.1.

If now we substitute the family (19.4) in the integral J, a function $J(b)$ of class C'' is obtained. We must have $J(b) \geqq J(0)$, if the set (19.4) is to form a minimal set for E_{12}. Hence we have

$$J'(0) = J_1(\eta) = \int_{x_1}^{x_2} \{f_{y_i}\eta_i + f_{y'_i}\eta'_i\}\, dx = 0 \tag{19.5}$$

along E_{12} for every set of variations η_i satisfying the conditions (19.3). It follows that the determinant

$$\begin{vmatrix} J_1(\eta_\nu) \\ J_2(\xi_\alpha, \eta_\nu) \end{vmatrix}$$

vanishes for every set of $m + 1$ variations $\eta_{i\nu}$ $(\nu = 1, \cdots, m + 1)$.[25] There exists therefore a set of $m + 1$ constants λ_0, λ_α not all zero such that

$$\lambda_0 J_1(\eta) + \lambda_\alpha J_2(\xi_\alpha, \eta) = 0$$

for every set of admissible variations η_i. If now we substitute for η_i the variations $\xi_{i\beta}$ it is found that

$$\lambda_\alpha J_2(\xi_\alpha, \xi_\beta) = 0,$$

and hence that the multipliers λ_α are all zero, since the set (19.1) forms a proper set. We may accordingly choose $\lambda_0 = 1$. The expression (19.5) must therefore hold for all admissible variations whatever. By the use of the fundamental lemma in the Calculus of Variations we obtain the following result:

THEOREM 19.1. *If the set* (19.1) *forms a proper minimal set for an admissible arc* E_{12}, *there exists a set of constants* c_i *not all zero such that the equations*

$$f_{y_i'} = \int_{x_1}^{x} f_{y_i}\, dx + c_i \tag{19.6}$$

hold at each point on E_{12}.

If the set (19.1) is a minimal set but not a proper minimal set, the function f in (19.6) must be replaced by a function of the form $F = \lambda_0 f + \lambda_\alpha f_\alpha$. This result follows from the theory of isoperimetric problems.

An admissible arc E_{12} will be said to satisfy the *condition of Weierstrass* if at each element (x, y, y') on it the inequality

$$E(x, y, y', Y') \geqq 0$$

for every admissible element $(x, y, Y') \neq (x, y, y')$, where

$$E = f(x, y, Y') - f(x, y, y') - (Y_i' - y_i')f_{y_i'}(x, y, y'). \tag{19.7}$$

We have the following result:

THEOREM 19.2. *Suppose that the conditions* (19.1) *form a proper minimal set for an admissible arc* E_{12}. *If* E_{12} *affords a strong minimum relatively to neighboring admissible arcs satisfying these conditions, then* E_{12} *satisfies the condition of Weierstrass.*

In order to obtain this result, let the point 3 be an arbitrary point on E_{12}, not a corner point or an end point of E_{12}. Let Y_{i3}' be an arbitrary direction and $Y_i(x)$ be an arbitrary curve of class C' passing through the point 3 and having $Y_i'(x_3) = Y_{i3}'$. The functions defining E_{12} will be denoted by $y_i(x)$. We now

[25] Cf. Bliss, *The problem of Lagrange in the Calculus of Variations*, American Journal of Mathematics, vol. 52 (1930), pp. 682, 692.

construct a one-parameter family of arcs $y_i(x, b)$ joining the end points of E_{12} by setting

$$y_i(x, b) = y_i(x) \qquad (x_1 \leqq x \leqq x_3),$$

$$y_i(x, b) = Y_i(x) \qquad (x_3 \leqq x \leqq b),$$

$$y_i(x, b) = y_i(x) + [Y_i(b) - y_i(b)]\frac{x - x_2}{b - x_2} \qquad (b \leqq x \leqq x_2).$$

This family is of the type usually used in the proof of the necessity of the Weierstrass condition for a minimum in the classical sense. It has the following important properties

$$\begin{aligned} &y_{ib}(x, x_3) \equiv 0 \quad (x_1 \leqq x \leqq x_3), \qquad y_{ib}(x_2, b) = 0, \\ &y_{ib}(x_3 + 0, x_3 + 0) = Y_i'(x_3) - y_i'(x_3). \end{aligned} \tag{19.8}$$

However, this family does not necessarily satisfy the conditions (19.1). In order to construct the one-parameter family of admissible arcs which satisfies these conditions, we substitute the functions

$$y_i(x, b) + a_\alpha \xi_{i\alpha}(x), \qquad y_{ix}(x, b) + a_\alpha \xi_{i\alpha}'(x) \tag{19.9}$$

for y_i, y_i' in $J_1(\xi_\alpha)$ and obtain a set of functions $J_{1\alpha}(a, b)$ having $J_{1\alpha}(0, x_3) = 0$. The functional determinant of these functions with respect to the variables $a_1, \cdots, a_m$ is precisely the determinant (19.2) and is accordingly different from zero. It follows that the equations $J_{1\alpha}(a, b) = 0$ have a unique solution $a_\alpha = A_\alpha(b)$ of class C' for $b \geqq x_3$ and sufficiently near $b = x_3$. Moreover $A_\alpha(x_3) = 0$. The one-parameter family

$$y_i = y_i(x, b) + A_\alpha(b)\xi_{i\alpha}(x) \tag{19.10}$$

therefore satisfies equations (19.1) for $b \geqq x_3$ and sufficiently near $b = x_3$, as was desired.

By substituting the functions (19.9) for y_i and y_i' in the integral J one obtains a function $J(a, b)$. Recalling that the family (19.10) satisfies the conditions (19.1), one sees that

$$J[A(b), b] \geqq J(x_3) \qquad (b \geqq x_3),$$

since the set (19.1) forms a minimal set for E_{12}. Hence we must have

$$(d/db)J[A(b), b]\big|_{b=0} = J_{a_\alpha}(0, x_3)A_\alpha'(0) + J_b(0, x_3) \geqq 0. \tag{19.11}$$

By differentiating the function $J(a, b)$ with respect to a_α it is found that

$$J_{a_\alpha}(0, x_3) = J_1(\xi_\alpha) = 0. \tag{19.12}$$

Moreover, if we write

$$J(0; b) = \int_{x_1}^{b} f\,dx + \int_{b}^{x_2} f\,dx,$$

by differentiating this expression with respect to b and setting $b = x_3$ we obtain from relations (19.8) and the usual integration by parts the relation

$$J_b(0, x_3) = f(x_3, y_3, Y_3') - f(x, y, y') - \int_{x_3}^{x_2} \{f_{y_i} y_{ib} + f_{y_i'} y_{ib}'\}\, dx$$
$$= E(x_3, y_3, y_3', Y_3').$$

Combining this result with the relations (19.12) and (19.11) we find that the theorem is true for all points on E_{12} except possibly at the corners and end points of E_{12}. That the theorem is also true at these points follows at once from continuity considerations.

An admissible arc E_{12} will be said to satisfy the *condition of Legendre* if at each element (x, y, y') on it the inequality

$$f_{y_i' y_k'} \pi_i \pi_k \geqq 0$$

holds for every set $(\pi) \neq (0)$. If this inequality holds with the equality sign excluded, then the *strengthened condition of Legendre* will be said to hold on E_{12}.

THEOREM 19.3. *If the conditions* (19.1) *form a proper minimal set for an admissible arc E_{12}, then the condition of Legendre holds on E_{12}. If E_{12} is nonsingular, then the strengthened condition of Legendre holds on E_{12}.*

To prove this, we note that according to the proof of the last theorem the function

$$\varphi(b) = E(x, y, y', y' + b\pi),$$

where the elements (x, y, y') are on E_{12} and (π) is arbitrary, must have a minimum at $b = 0$. It follows that $\varphi'(0) = 0, \varphi''(0) = f_{y_i' y_k'} \pi_i \pi_k \geqq 0$, as was to be proved. This simple proof has been used by Morse.

Theorems 19.3 and 19.4 give a new and broader significance to the conditions of Weierstrass and Legendre in the sense that they are not only necessary conditions for a minimum in the classical sense but also in the more general sense here defined.

We have the further result

THEOREM 19.4. *If the set* (19.1) *forms a proper minimal set for an extremal arc E_{12}, the inequality $J_2(\eta) \geqq 0$ is true for every admissible variation $\eta_i(x)$ satisfying the m conditions $J_2(\xi_\sigma, \eta) = 0$ along E_{12}.*

This result is obtained by differentiating the function $J(b)$ appearing in the proof of Theorem 19.1 twice for b and noting that $J''(0) = J_2(\eta)$ must be positive or zero if the set (19.1) is to form a minimal set for E_{12}.

20. Mayer fields and an analogue of Hahn's Lemma. Consider now an n-parameter family of extremals

$$y_i = y_i(x, a_1, \cdots, a_n) \tag{20.1}$$

whose functions $y_i(x, a)$, $y_{ix}(x, a)$ are of class C'' in a region of points (x, a) with x on $\bar{x}_1 \leqq x \leqq \bar{x}_i$ and a_i in A. Suppose further that the determinant $| y_{ia_k} |$ is

different from zero within this region and that this region has been chosen so small that the equations (20.1) have unique solutions $a_i = a_i(x, y)$ of class C'' for all (x, y) in a region $\mathfrak{F}$ in xy-space. The family (20.1) is then said to *simply cover* the region $\mathfrak{F}$. Let

$$p_i(x, y) = y_{ix}[x, a(x, y)].$$

The region $\mathfrak{F}$ will be called a *Mayer field* if the Hilbert integral

$$J^* = \int f(x, y, p)\, dx + (dy_i - p_i\, dx) f_{y_i'}(x, y, p)$$

with $p_i = p_i(x, y)$ is independent of the path in $\mathfrak{F}$. The functions $p_i(x, y)$ are called the *slope functions* of the field and the solutions (20.1) of the equations $y_i' = p_i(x, y)$ are called the *extremals of the field* $\mathfrak{F}$. Along an extremal of the field we have clearly $J^* = J$.

The following theorem is well-known. In this theorem the function $E(x, y, p, y')$ is the Weierstrass E-function defined by the equation (19.7).

THEOREM 20.1. *If E_{12} is an extremal of a field $\mathfrak{F}$ at each point of which the condition*

$$E[x, y, p(x, y), y'] > 0$$

holds for every admissible set $(x, y, y') \neq (x, y, p)$, then the inequality $J(C_{12}) > J(E_{12})$ is true for every admissible arc C_{12} in $\mathfrak{F}$ joining the end points of E_{12} but not identical with E_{12}.

This result follows in the usual manner by the use of the Weierstrass integral formula

$$J(C_{12}) - J(E_{12}) = \int_{C_{12}} E(x, y, p, y')\, dx ,$$

which is readily established by the use of the equations

$$J(E_{12}) = J^*(E_{12}) = J^*(C_{12}).$$

We shall need the following lemma:

LEMMA 20.1. *If the Hilbert integral J^* formed for the family* (20.1) *is independent of the path on the hyperplane $x = x_0$, the family* (20.1) *defines a Mayer field over every region $\mathfrak{F}$ which it simply covers.*

This result is well-known and is due to Hilbert.[26]

A lemma of a different type is the following one:

LEMMA 20.2. *If a non-singular extremal E_{12} has on it no point 3 conjugate to 1, there exists a conjugate system of accessory extremals η_{ik} for E_{12} having its determinant different from zero on E_{12}.*

This result and its proof are also well known. One needs only to take the

[26] Göttinger Nachrichten, (1905), p. 165; Math. Annalen, 62 (1906), p. 356; see also Bliss, loc. cit., p. 733.

conjugate system having $\eta_{ik} = 0$ at a value x_0 to the left of x_1 and sufficiently near x_1.

Consider now a proper set of m natural isoperimetric conditions

$$J_1(\xi_\alpha) = \int_{x_1}^{x_2} f_\alpha \, dx = 0 \qquad (\alpha = 1, \cdots, m), \tag{20.2}$$

and let $F = f + \lambda_\alpha f_\alpha$. The Weierstrass E-function formed for the function F will be denoted by $E_\lambda(x, y, y', Y')$. It is precisely the E-function for the integral

$$J_\lambda = \int_{x_1}^{x_2} F(x, y, y', \lambda) \, dx.$$

An extremal arc E_{12} for our original integral J will be said to satisfy the *strengthened condition of Weierstrass relative to the conditions* (20.2) if at each element (x, y, y') in a neighborhood $\mathfrak{R}$ of those on E_{12} the inequality

$$E_\lambda(x, y, y', Y') > 0$$

holds for every admissible set $(x, y, Y') \neq (x, y, y')$ and for every set of constants (λ_α) in a neighborhood Λ of $(\lambda) = (0)$. If there are no conditions of the type (20.2), this condition becomes the usual strengthened condition of Weierstrass.

We have the following analogue of Hahn's Lemma.[27] The proof here given is like that of Bliss.[28]

LEMMA 20.3. *Let E_{12} be an extremal arc having on it no point conjugate to its initial end point* 1. *If the strengthened conditions of Weierstrass and Legendre hold on E_{12}, there exists a neighborhood $\mathfrak{F}$ of E_{12} in xy-space and a neighborhood N of the end points of E_{12} in $(x_1y_1x_2y_2)$-space such that every isoperimetric extremal E_λ in $\mathfrak{F}$ with end points in N and having its multipliers* (λ) *sufficiently near* $(\lambda) = (0)$ *affords a proper strong minimum to J_λ relative to admissible arcs in $\mathfrak{F}$ joining the end points of E_λ.*

To prove this let η_{ik} be a conjugate system of accessory extremals for E_{12} having its determinant $\eta_{ik}(x)$ different from zero on x_1x_2. Such a conjugate system surely exists by virtue of Lemma 20.2, since in this case there is no point on E_{12} conjugate to the point 1. Let ζ_{ik} be the values of $\omega_{\eta_i'}$ along η_{ik} on E_{12}. We may suppose that the functions η_{ik} have been chosen so that $\eta_{ik} = \delta_{ik}$ at $x = x_1$.

Consider now the $(2n + m)$-parameter family of isoperimetric extremals

$$y_i = Y_i(x, a_1, \cdots, a_n, b_1, \cdots, b_n, \lambda_1, \cdots, \lambda_m)$$

described in §18. The constants a_i, b_i may be chosen to be the values of y_i,

[27] Hahn, *Über Variationsprobleme mit variablen Endpunkten*, Monatshefte für Mathematik und Physik, vol. 22 (1911), pp. 127–136.

[28] Bliss, *The problem of Bolza in the Calculus of Variations*, Annals of Mathematics, vol. 33 (1932), pp. 267–270.

and $z_i = f_{y_i'} + \lambda_\alpha f_{\alpha y'}$ at $x = x_1$. Let $a_{i0}, b_{i0}, \lambda_\alpha = 0$ be the values of (a, b, λ) belonging to E_{12} and let

$$2w(a) = \zeta_{ik}(x_1)(a_i - a_{i0})(a_k - a_{k0}).$$

The family

$$y_i = y_i(x, a, b, \lambda) = Y_i(x, a, b + w_a, \lambda) \tag{20.3}$$

contains E_{12} for $a_i = a_{i0}, b_i = b_{i0}, \lambda_\alpha = 0$, and has

$$y_{ia_k}(x_1, a, b, \lambda) = \delta_{ik} = \eta_{ik}(x_1), \qquad z_{ia_k}(x_1, a, b, \lambda) = \zeta_{ik}(x_1), \tag{20.4}$$

where $z_i = F_{y_i'}$. Along E_{12} the functions y_{ia_k} form a set of n accessory extremals for which z_{ia_k} represents the values of $\omega_{\eta_i'}$. Hence we must have $y_{ia_k} = \eta_{ik}$ along E_{12} since they have the same initial conditions at $x = x_1$, by virtue of the equations (20.4). The determinant $| y_{ia_k} |$ is accordingly different from zero along E_{12} and hence also for all values of (x, y, b, λ) near those on E_{12}. The equations (20.3) therefore have a unique solution $a_i = a_i(x, y, b, \lambda)$ of class C'' for values (x, y, b, λ) in a neighborhood S of those on E_{12}. Let $\mathfrak{F}$ be the projection of this neighborhood in xy-space. For fixed values of (b, λ) belonging to a set (x, y, b, λ) in S the n-parameter family (20.3) defines a Mayer field with slope functions

$$p_i(x, y, b, \lambda) = y_{ix}[x, a(x, y, b, \lambda), b, \lambda]$$

over the region $\mathfrak{F}$ in xy-space. This follows from Lemma 20.1, the Hilbert integral J^* being independent of the path on the hyperplane $x = x_1$ since on this hyperplane we have

$$J^* = \int d(w + b_i a_i),$$

as one readily verifies. Moreover, if the region S is diminished still further, then by virtue of the strengthened condition of Weierstrass for E_{12} the inequality

$$E_\lambda[x, y, p(x, y, b, \lambda), y'] > 0$$

will hold for every set (x, y, p, λ) in S and for every admissible set $(x, y, y') \neq (x, y, p)$. It follows from Theorem 20.1 that every extremal of the family (20.3) with elements (x, y, b, λ) in S and $a_i = a_i(x, y, b, \lambda)$ will afford a proper minimum to the integral J relative to admissible arcs in $\mathfrak{F}$ joining its end points.

The proof of Lemma 20.3 will be complete if we show that every isoperimetric extremal with end points $(x_1, y_{i1}, x_2, y_{i2})$ in a neighborhood N of those of E_{12} and with multipliers (λ) in a neighborhood Λ of $(\lambda) = (0)$ is an extremal of the family (20.3) with elements (x, y, b, λ) in S and $a_i = a_i(x, y, b, \lambda)$. This result follows at once. For, since the end points of E_{12} are not conjugate, every pair of points (x_1, y_1), (x_2, y_2) in N can be joined by an extremal E of the family (20.3). If N and Λ are taken sufficiently small, the elements (x, y, b, λ) belonging to E will lie in S. The constants a_i must accordingly be given by the functions $a_i(x, y, b, \lambda)$, since these solutions are unique. Lemma 20.3 is now proved.

21. The existence of minimal sets. We are now in position to establish the following basic result:

THEOREM 21.1. *Let s be the sum of the orders of the conjugate points of the point 1 on an extremal arc E_{12}. If E_{12} satisfies the strengthened conditions of Weierstrass and Legendre, and the point 2 is not conjugate to the point 1, then there exists a set of s natural isoperimetric conditions*

$$J_1(\xi_\alpha) = 0 \qquad (\alpha = 1, \cdots, s)$$

such that E_{12} affords a proper strong minimum to the integral J relative to neighboring admissible arcs joining its end points and satisfying the conditions (20.1). *Moreover, no proper subset of these conditions has this property.*

As a first step in the proof of this theorem let

$$t_0 = x_1 < t_1 < \cdots < t_q < t_{q+1} = x_2$$

be a set of points on the x-axis such that there are no pairs of conjugate points on the sub-arc of E_{12} defined by the successive points determined by the values t_r and t_{r+1}. Let $\xi_{i\alpha}(x)$ be a set of s functions of class C^3 such that the equations

$$J_2(\xi_\alpha, \eta) = 0 \qquad (\alpha = 1, \cdots, s) \tag{21.1}$$

form a proper minimal set of natural isoperimetric conditions for the second variation $J_2(\eta)$ along E_{12}. The number of these conditions is equal to the sum of the orders of the conjugate points of the point 1 on E_{12}, by Theorem 5.1. Moreover, according to the corollary of Theorem 3.1 we have

$$J_2(\xi_\alpha, \xi_\beta)a_\alpha a_\beta < 0 \qquad (\alpha, \beta = 1, \cdots, s)$$

for every set of constants $(a) \neq (0)$. Let $\eta_{i\alpha}$ be a set of broken accessory extremals having

$$\eta_{i\alpha}(t_r) = \xi_{i\alpha}(t_r) \qquad (r = 0, 1, \cdots, q + 1). \tag{21.2}$$

Since each extremal segment of $\eta_{i\alpha}$ is a minimizing arc for $J_2(\eta)$ we have

$$J_2(\eta_\alpha, \eta_\beta)a_\alpha a_\beta \leqq J_2(\xi_\alpha, \xi_\beta)a_\alpha a_\beta < 0. \tag{21.3}$$

The equality holds only in case the a's are all zero, since otherwise the functions $\xi_{i\alpha}a_\alpha$, $\eta_{i\alpha}a_\alpha$ would be identical. This, however, is impossible since the first of these arcs has no corners. An important consequence of this fact is that the determinant

$$| J_2(\eta_\alpha, \eta_\beta) - J_2(\xi_\alpha, \xi_\beta) | \tag{21.4}$$

is different from zero along E_{12}.

As a second step in the proof of Theorem 21.1 let

$$J_1(\xi_\alpha) = \int_{x_1}^{x_2} f_\alpha \, dx = \int_{x_1}^{x_2} \{f_{y_i}\xi_{i\alpha} + f_{y_i'}\xi'_{i\alpha}\} \, dx = 0 \tag{21.5}$$

be the natural isoperimetric conditions for E_{12} determined by the functions $\xi_{i\alpha}$ just described. The Euler equations of the integral J subject to these conditions are the equations (18.3). Let

$$y_i = y_i(x, b_{11}, \cdots, b_{qn}, \lambda_1, \cdots, \lambda_s) \qquad (x_1 \leqq x \leqq x_2) \tag{21.6}$$

be a $(qn + s)$-parameter family of broken isoperimetric extremals having its corners on the hyperplanes

$$x = t_1, \qquad x = t_2, \cdots, x = t_q$$

and containing E_{12} for the special values

$$b_{ir} = b_{ir0}, \qquad \lambda_\alpha = 0 \qquad (i = 1, \cdots, n; r = 1, \cdots, q; \alpha = 1, \cdots, s).$$

The parameter b_{ir} can be chosen to be the values of y_i on the hyperplane $x = t_r$. We have accordingly the equations

$$y_i(x_1, b, \lambda) = y_{i1}, \qquad y_i(t_r, b, \lambda) = b_{ir}, \qquad y_i(x_2, b, \lambda) = y_{i2}.$$

If now we set $u_{i\alpha} = y_{i\lambda_\alpha}(x, b_0, 0)$, then by differentiating these equations with respect to λ_α it is seen that

$$u_{i\alpha}(x_1) = 0, \qquad u_{i\alpha}(t_r) = 0, \qquad u_{i\alpha}(x_2) = 0. \tag{21.7}$$

Upon substituting the family (21.6) in the Euler equations (18.3) and differentiating with respect to λ_α again we obtain by setting $(b, \lambda) = (b_0, 0)$ the equations

$$L_i(u_\alpha + \xi_\alpha) = 0.$$

It follows that the variations $u_{i\alpha} + \xi_{i\alpha}$ form a set of s broken accessory extremals $\eta_{i\alpha} = u_{i\alpha} + \xi_{i\alpha}$. By virtue of the equations (21.7) these extremals satisfy the conditions (21.2) and are accordingly identical with the extremals $\eta_{i\alpha}$ described in the last paragraph. The determinant

$$| J_2(\xi_\alpha, u_\beta) | \qquad (\alpha, \beta = 1, \cdots, s) \tag{21.8}$$

can now be shown to be different from zero on the arc E_{12}. For, if we substitute for $u_{i\beta}$ its values $\eta_{i\beta} - \xi_{i\beta}$ it is found that

$$J_2(\xi_\alpha, u_\beta) = J_2(\xi_\alpha, \eta_\beta) - J_2(\xi_\alpha, \xi_\beta).$$

Moreover, $J_2(\xi_\alpha, \eta_\beta) = J_2(\eta_\alpha, \eta_\beta)$, as can be seen by the usual integration by parts with the help of equations (21.2). The determinant (21.8) is therefore equal to the determinant (21.4) and hence must be different from zero, as was to be proved.

As a third step in the proof of Theorem 21.1 we substitute the family (21.6) in the functions $J_1(\xi_\alpha)$ and obtain a set of s functions $J_{1\alpha}(b, \lambda)$ having $J_{1\alpha}(b_0, 0) = 0$. The determinant of the derivatives of these functions with respect to the variables $\lambda_1, \cdots, \lambda_m$ at $(b, \lambda) = (b_0, 0)$ is equal to the determinant (21.8) and is therefore different from zero. The equations $J_{1\alpha}(b, \lambda) = 0$

accordingly have solutions $\lambda_\alpha = \lambda_\alpha(b_{11}, \cdots, b_{qn})$ of class C'' such that $\lambda_\alpha(b_0) = 0$. When the functions $\lambda_\alpha(b)$ are substituted in the equations (21.6) a new qn-parameter family

$$(21.9) \qquad y_i = y_i(x, b_{11}, \cdots, b_{qn}), \qquad \lambda_\alpha = \lambda_\alpha(b_{11}, \cdots, b_{qn}) \qquad (x_1 \leqq x \leqq x_2)$$

of broken extremals is obtained which satisfies the conditions (21.5) and contains E_{12} for $b_{ir} = b_{ir0}$. For this new family the parameters b_{ir} still represent the values of y_i on the hyperplane $x = t_r$. It follows that the variations

$$\delta y_i = y_{ib_{jr}}(x, b_0)\, db_{jr} \qquad (j, r \text{ both summed})$$

are not all identically zero if the constants db_{jr} are not all zero. Since the family (21.9) satisfies the conditions (21.5), it is clear that the variations $\eta_i = \delta y_i$ must satisfy the conditions (21.1). Hence we have $J_2(\delta y) > 0$ along E_{12}, by Theorem 5.1, since the point 2 is not conjugate to 1. An important consequence of this fact is that the functional J takes a proper minimum on E_{12} relative to the family (21.9). To prove this we need only substitute the family (21.9) in the integral J and note that for the function $J(b)$ thus obtained we have $dJ = 0$, $d^2J = J_2(\delta y) > 0$, $(\delta y) \not\equiv (0)$, at $b_{ir} = b_{ir0}$ and hence $J(b) > J(b_0)$ for $(b) \neq (b_0)$.

We are now in position to complete the proof of Theorem 21.1. To do so let $\mathfrak{F}'$ be a neighborhood of E_{12} so small that each extremal sub-arc of the family (21.9) with ends on successive hyperplanes $x = t_r$, $x = t_{r+1}$ affords a strong minimum to the integral

$$J_\lambda = \int_{x_1}^{x_2} \{f + \lambda_\alpha f_\alpha\}\, dx$$

relative to admissible arcs in $\mathfrak{F}'$ joining its end points. This is possible by virtue of Lemma 20.3. Let $\mathfrak{F}$ be a neighborhood of E_{12} interior to $\mathfrak{F}'$ such that every admissible arc C joining the end points cuts the hyperplanes $x = t_r$ in points (t_r, b_{ir}) whose coördinates b_{ir} determine an extremal E_b of the family (21.9) lying in $\mathfrak{F}'$. By Lemma 20.3 we have $J_\lambda(C) \geqq J_\lambda(E_b)$, the equality holding only in case $C \equiv E_b$. Suppose now that the arc C satisfies the conditions (21.5). We then have $J(C) \geqq J(E_b)$, since the arc E_b also satisfies these conditions. If necessary, we may diminish the region $\mathfrak{F}$ still further so that $J(E_b) \geqq J(E_{12})$, as described in the last paragraph. We then have $J(C) > J(E_{12})$ unless $C \equiv E_{12}$. The proof of Theorem 21.1 is now complete.

By the use of Taylor's expansion with integral remainder it is found that the strengthened condition of Legendre implies the strengthened condition of Weierstrass, if we restrict the admissible sets (x, y, y') to lie in a sufficiently small neighborhood $\mathfrak{R}_1$ of those on E_{12}. Hence we have the following result:

Corollary 1. *If the phrase "strong minimum" is replaced by "weak minimum," the conclusions of Theorem 21.1 are still true if the extremal E_{12} satisfies the strengthened condition of Legendre but not necessarily that of Weierstrass, at least if the final end point 2 is not conjugate to the initial point 1 on E_{12}.*

We have further

COROLLARY 2. *If an extremal E_{12} satisfies the strengthened condition of Legendre and the point 2 is not conjugate to 1, the number of conditions in a minimal set of natural isoperimetric conditions for E_{12} is always the same and is equal to the sum of the orders of the conjugate points of the point 1 on E_{12}.*

With the help of the corollary to Theorem 3.1 we obtain the further result:

COROLLARY 3. *If the strengthened condition of Legendre holds on an extremal arc E_{12} whose end points are not conjugate, and if the set (21.5) forms a proper minimal set for E_{12}, the inequality*

$$J_2(\xi_\alpha, \xi_\beta)a_\alpha a_\beta < 0 \qquad (\alpha, \beta = 1, \cdots, s)$$

holds on E_{12} for every set of constants $(a) \neq (0)$.

22. **The variable end point case.** Consider now the functional

$$J = g[y(x_1), y(x_2)] + \int_{x_1}^{x_2} f(x, y, y')\, dx$$

subject to a set of conditions of the form

$$\psi_\mu[y(x_1), y(x_2)] = 0 \qquad (\mu = 1, \cdots, p \leqq 2n).$$

The values x_1, x_2 are, of course, assumed to be constants. The functions $g(y_1, y_2)$, $\psi_\mu(y_1, y_2)$ are assumed to be of class C'' in a neighborhood of the end values of the extremal E_{12} in question and the matrix

$$\| \psi_{\mu y_{i1}} \quad \psi_{\mu y_{i2}} \|$$

is assumed to have rank p. An arc C of class D' with elements (x, y, y') in the region R described in §17 will be called an *admissible arc* for the variable end point case if it satisfies the conditions $\psi_\mu = 0$.

If we set $G = g + l_\mu \psi_\mu$, the transversality conditions for the problems here considered are given by the equations

$$(22.1) \qquad f_{y_i'}(x_1) = G_{y_{i1}}, \qquad f_{y_i'}(x_2) = -G_{y_{i2}}.$$

Along an extremal E_{12} satisfying the end conditions $\psi_\mu = 0$ and the transversality conditions (22.1) with a set of constants l_μ the second variation is expressible in the form

$$(22.2) \qquad J_2(\eta) = 2q[\eta(x_1), \eta(x_2)] + \int_{x_1}^{x_2} 2\omega(x, \eta, \eta')\, dx,$$

where 2ω is defined as before, and

$$2q = G_{y_{i1}y_{j1}}\eta_{i1}\eta_{j1} + 2G_{y_{i1}y_{j2}}\eta_{i1}\eta_{j2} + G_{y_{i2}y_{j2}}\eta_{i2}\eta_{j2}.$$

To prove this let

$$y_i = y_i(x, b) \qquad (x_1 \leqq x \leqq x_2)$$

be a one-parameter of admissible arcs containing E_{12} for $b = 0$ and having suitable continuity properties. Substituting this expression in the functional J we obtain a function $J(b)$. This function is expressible in the form

$$J(b) = G[y(x_1, b), y(x_2, b), l] + \int_{x_1}^{x_2} f[x, y(x, b), y'(x, b)]\, dx,$$

where l_μ are the constants belonging to E_{12}. Differentiating twice with respect to b and setting $b = 0$, $\eta_i = y_{ib}(x, 0)$ one obtains the expression (22.2) by the usual integration by parts and by the use of equations (22.1), as one readily verifies.

The equations of variation of the end conditions $\psi_\mu = 0$ on E_{12} are the equations

$$\Psi_\mu(y) = \psi_{\mu y_{i1}} \eta_i(x_1) + \psi_{\mu y_{i2}} \eta_i(x_2) = 0 . \tag{22.3}$$

It follows that the functional (22.2) subject to the conditions (22.3) determines a problem of the type described in §§7–10. If functional $J_2(\eta)$ is non-degenerate on E_{12}, then the extremal E_{12} will be said to be *non-degenerate*. By the *order of concavity* of E_{12} will be meant the order of concavity of the functional $J_2(\eta)$ relative to the end conditions.

Let $\xi_i[x, y, y_1, y_2]$ be a set of functions of class C^4 satisfying the conditions

$$\psi_{\mu y_{i1}}\xi_{i1} + \psi_{\mu y_{i2}}\xi_{i2} = 0 ,$$

where

$$\xi_{i1} = \xi[x_1, y(x_1), y(x_1), y_i(x_2)],\ \xi_{i2} = \xi_i[x_2, y(x_2), y(x_1), y(x_2)] .$$

A condition of the form

$$J_1(\xi) = g_{y_{i1}}\xi_{i1} + g_{y_{i2}}\xi_{i2} + \int_{x_1}^{x_2} \{f_{y_i}\xi_i + f_{y_i'}\xi_i'\}\, dx = 0$$

where $\xi_i' = \xi_{ix} + \xi_{iy_k} y_k'$ will be called a *natural isoperimetric condition*. It is satisfied by every extremal arc satisfying the conditions $\psi_\mu = 0$ and the transversality conditions (22.1).

Let E_{12} be an extremal arc satisfying the end conditions $\psi_\mu = 0$ and the transversality conditions (22.1). A set of m natural isoperimetric conditions

$$J_1(\xi_\alpha) = 0 \qquad (\alpha = 1, \cdots, m) \tag{22.4}$$

will be called a *minimal set* for E_{12} if the arc E_{12} affords at least a weak relative minimum to the functional J relative to neighboring arcs satisfying the conditions (22.4), and if no proper sub-set of these conditions has this property. Such a set will be said to form a *proper set* for E_{12} if the determinant

$$| J_2(\xi_\alpha, \xi_\beta) |$$

is different from zero on E_{12}. By an argument similar to that used in §19 it can be shown that if the set (22.4) is a proper set for E_{12}, there are infinitely many admissible arcs satisfying the conditions (22.4) in every neighborhood of E_{12}.

THEOREM 22.1. *Suppose the conditions* (22.4) *form a proper minimal set for the extremal arc* E_{12}. *If* E_{12} *affords a strong minimum of* J *relative to neighboring admissible arcs satisfying these conditions* (22.4), *the condition of Weierstrass holds on* E_{12}.

The proof is similar to that of Theorem 19.2. Let $\eta_{i\gamma}(x)$ be a set of $(m + p)$ variations such that the determinant

$$\begin{vmatrix} J_2(\xi_\alpha, \eta_\gamma) \\ \psi_\mu(\eta_\gamma) \end{vmatrix} \qquad (\gamma = 1, \cdots, m + p) \tag{22.5}$$

is different from zero on E_{12}, and set

$$y_i(x, a, b) = y_i(x, b) + a_\gamma \eta_{i\gamma}(x) \tag{22.6}$$

where $y_i(x, b)$ is the family used in equations (19.9). When this family is substituted in the expressions $J_1(\xi_\alpha)$, ψ_μ, a set of $m + p$ functions $J_{1\alpha}(a, b)$, $\psi_\mu(a, b)$ is obtained having $J_{1\alpha}(0, x_3) = 0$, $\psi_\mu(0, x_3) = 0$. Moreover, the functional determinant of $J_{1\alpha}(a, b)$, $\psi_\mu(a, b)$ with respect to $a_1, \cdots, a_{m+p}$ at $(a, b) = (0, x_3)$ is the determinant (22.5) and is accordingly different from zero. The equations

$$J_1(a, b) = 0, \qquad \psi_\mu(a, b) = 0$$

therefore have solutions $a_\gamma = a_\gamma(b)$ of class C^1 with $a_\gamma(x_3) = 0$. The family

$$y_i = y_i(x, b) + a_\gamma(b)\eta_{i\gamma}(x) \qquad (x_1 \leqq x \leqq x_2) \tag{22.7}$$

satisfies the conditions (22.4) and $\psi_\mu = 0$ and contains E_{12} for $b = x_3$. The remainder of the proof is now like that of Theorem 19.2 if we replace the families (19.9), (19.10) by (22.6), (22.7) respectively.

THEOREM 22.2. *If the set* (22.4) *forms a proper minimal set for* E_{12}, *the condition of Legendre holds on* E_{12}. *If* E_{12} *is non-singular, the strengthened condition of Legendre holds.*

The analogue of Theorem 19.4 is the following

THEOREM 22.3. *If the set* (22.4) *forms a proper minimal set for* E_{12}, *the inequality* $J_2(\eta) \geqq 0$ *is true for every set of admissible variations satisfying the condition* $\Psi_\mu = 0$, $J_2(\xi_\alpha, \eta) = 0$ *along* E_{12}.

If our end conditions are of the special form

$$\psi_{\mu 1}[y(x_1)] = 0, \qquad \psi_{\mu 2}[y(x_2)] = 0, \tag{22.8}$$

the functions ξ_i used in the definition of natural isoperimetric conditions can be taken in the form $\xi_i[x, y(x)]$. The following theorem can then be established by arguments similar to those used in the proof of Theorem 22.1. We assume that E_{12} is not tangent to the end manifolds and that the end points of E_{12} are the only pairs of points on E_{12} satisfying these end conditions.

THEOREM 22.4. *Let* E_{12} *be a non-degenerate extremal satisfying the end conditions* (22.8) *and the transversality conditions* (22.1). *Let* r *be the order of con-*

cavity of E_{12} and s the sum of the orders of the conjugate points of 1 *on E_{12} between* 1 *and* 2. *If E_{12} satisfies the strengthened conditions of Weierstrass and Legendre, there exists a proper set of $m = r + s$ natural isoperimetric conditions* (22.4), *such that E_{12} affords a proper strong minimum for the integral J relative to neighboring admissible arcs satisfying these conditions. Moreover, no proper sub-set of these conditions has this property.*

It is interesting to note that Theorem 22.5 holds even if the point 2 is conjugate to 1. Our only restriction is that the arc E_{12} be non-degenerate.

COROLLARY 1. *The conclusions of Theorem* 22.5 *are still true if we replace the phrase "strong minimum" by "weak minimum" and omit the condition of Weierstrass.*

COROLLARY 2. *If the strengthened condition of Legendre holds on a non-degenerate extremal E_{12} which satisfies the transversality conditions* (22.1), *the number of conditions in a minimal set is always the same and is equal to the order of concavity plus the sum of the orders of the conjugate points of the point* 1 *on E_{12} between* 1 *and* 2.

The analogue of Corollary 3 of Theorem 21.1 is also true, as one readily verifies.

IV

The general problem in parametric form

In §§23–27 we shall be concerned with a functional of the form

$$J = g[x(t_1), x(t_2)] + \int_{t_1}^{t_2} f(x, \dot{x})\, dt$$

defined on an $(n + 1)$-dimensional Riemannian manifold $\Re$ with local coördinates of the form $(x^0, x^1, \cdots, x^n)$. Our end conditions will be taken in the form

$$\psi_\mu[x(t_1), x(t_2)] = 0 \qquad (\mu = 1, \cdots, p \leqq 2\,n + 2).$$

In our sufficiency theorems we shall restrict ourselves to the *separated end point case*, that is, to the case in which the end conditions are of the form

$$\psi_{\mu_1}[x(t_1)] = 0, \qquad\qquad \psi_{\mu_2}[x(t_2)] = 0.$$

In this part we shall show that if a non-degenerate extremal arc E_{12} satisfies these end conditions, the associated transversality condition and the *usual* strengthened conditions of Weierstrass and Legendre, there exists a set of m natural isoperimetric conditions such that E_{12} affords a strong proper minimum relative to neighboring admissible arcs satisfying these isoperimetric conditions and the end conditions. Moreover, there is a smallest set of natural isoperimetric conditions and the number of conditions in such a set is equal to the order of concavity of E_{12} plus the sum of the orders of the conjugate points of the point 1 on E_{12} between 1 and 2. It should be noted that this result is true even if the end points of E_{12} are conjugate to each other, provided that E_{12} is

non-degenerate. In the fixed end point case E_{12} is clearly degenerate if its end points are conjugate. Of particular interest is Theorem 26.1, below, which gives an extended sufficiency theorem for the fixed end point case. The results of this part are obtained without the use of the classical isoperimetric theory. In this case we do not need to modify the strengthened condition of Weierstrass as was done in the non-parametric case.

23. **Hypotheses and preliminary definitions.** The function $f(x, \dot{x})$ is assumed to be of class C^4 for local coördinates (x) on the Riemannian manifold $\mathfrak{R}$ and for all vectors $(\dot{x}) \neq (0)$. The function $f(x, \dot{x})$ is defined so as to be invariant under transformations of the form

$$y^i = y^i(x^0, x^1, \cdots, x^n), \qquad \dot{y}^i = \frac{\partial y^i}{\partial x^k}\dot{x}^k \qquad (i, k = 0, 1, \cdots, n),$$

where the functions $y^i(x)$ are of class C^5. We assume also that f is positively homogeneous of order one in the variables $(\dot{x})$.

Let $\mathfrak{R}^2$ be the product manifold $\mathfrak{R} \times \mathfrak{R}$. Its local coördinates

$$(x_1^0, \cdots, x_1^n, x_2^0, \cdots, x_2^n)$$

are composed of the coördinates (x_1) and (x_2) of its image in $\mathfrak{R}$ and are to be transformed accordingly. The functions $g(x_1, x_2)$ and $\psi_\mu(x_1, x_2)$ are assumed to be of class C^3 on $\mathfrak{R}^2$.

A *regular arc* is a sensed arc defined locally by equations of the form

$$x^i = x^i(t) \qquad (t_1 \leqq t \leqq t_2)$$

of class C' and having $(\dot{x}) \neq (0)$. A neighborhood of a regular arc E_{12} which does not intersect itself can be imbedded in a single admissible coördinate system (x) with $\dot{x}^0 > 0$ along E_{12}. It will be understood throughout this part that such a coördinate system has been chosen. Our results are, of course, independent of the coördinate system used.

By an *admissible arc* will be meant a continuous sensed arc which is composed of a finite number of regular sub-arcs and which satisfies the conditions $\psi_\mu = 0$. At times it will be convenient to call an arc admissible even if the particular set of end conditions $\psi_\mu = 0$ here given are not satisfied. In such a case an explicit description of the end conditions used will be given. Unless otherwise expressly stated, it will be assumed that the matrix

$$\| \psi_{\mu x_1^i} \quad \psi_{\mu x_2^i} \|$$

has rank p on the arc E_{12} in question.

The Euler equations associated with the functional J are the equations

$$P_i = f_{x^i} - (d/dt)f_{\dot{x}^i} = 0 \qquad (i = 0, 1, \cdots, n). \tag{23.1}$$

A solution of these equations of class C'' is called an *extremal arc*. Such an arc will be said to be *non-singular* in case the matrix

$$\| f_{\dot{x}^i\dot{x}^k} \| \qquad (i, k = 0, 1, \cdots, n) \tag{23.2}$$

has rank n. For a non-singular extremal arc the parameter t can be chosen so that the functions $x^i(t)$ defining E_{12} are of class C^4.

The transversality conditions associated with the functional J are the equations

$$f_{\dot{x}^i}(x_1, \dot{x}_1) = G_{x_1^i}, \qquad f_{\dot{x}^i}(x_2, \dot{x}_2) = -G_{x_2^i}, \tag{23.3}$$

where $G = g + l_\mu \psi_\mu$.

24. **The accessory problem.** Along an extremal arc E_{12} satisfying the end conditions $\psi_\mu = 0$ and the transversality conditions (23.3) the second variation is always expressible in the form

$$J_2(\eta) = 2q(\eta_1, \eta_2) + \int_{t_1}^{t_2} 2\omega(\eta, \dot{\eta})\, dt,$$

where

$$2q = G_{x_1^i x_1^j}\eta_1^i\eta_1^j + 2G_{x_1^i x_2^j}\eta_1^i\eta_2^j + G_{x_2^i x_2^j}\eta_2^i\eta_2^j,$$

$$2\omega = f_{x^i x^j}\eta^i\eta^j + 2f_{x^i\dot{x}^j}\eta^i\dot{\eta}^j + f_{\dot{x}^i\dot{x}^j}\dot{\eta}^i\dot{\eta}^j.$$

This result follows at once if we consider our problem as a non-parametric problem in (tx)-space with t_1 and t_2 fixed. The variations $\eta^i(t)$ are assumed to be of class D' and to satisfy the conditions

$$\Psi_\mu(\eta) = \psi_{\mu x_1^i}\eta^i(t_1) + \psi_{\mu x_2^i}\eta^i(t_2) = 0 \qquad (\mu = 1, \cdots, p).$$

Such variations will be termed *admissible variations*.

The problem determined by the functional $J_2(\eta)$ subject to the conditions $\Psi_\mu = 0$ will be called the *accessory problem*. It is of the type described in §§7–10. We define, as before, a *natural isoperimetric* condition to be a condition of the form

$$J_2(\xi, \eta) = Q(\xi, \eta) + \int_{t_1}^{t_2} \Omega(\xi, \eta)\, dt = 0,$$

where

$$Q(\xi, \eta) = \xi_1^i q_{\eta_1^i} + \xi_2^i q_{\eta_2^i} = Q(\eta, \xi),$$

$$\Omega(\xi, \eta) = \xi^i\omega_{\eta_i} + \dot{\xi}^i\omega_{\dot{\eta}_i} = \Omega(\eta, \xi),$$

and the functions $\xi^i(t)$ define admissible variations of class C''. We define minimal sets, proper sets, type number, etc., as in §8. However, Theorem 10.1 is not applicable in this case since the determinant (23.2) is now identically zero.

The accessory equations are the equations

$$L_i(\eta) = \omega_{\eta_i} - (d/dt)\omega_{\dot{\eta}_i} = 0. \tag{24.1}$$

A solution of class C'' of these equations will be called an *accessory extremal*. It is well known that if $\rho(t)$ is of class D', then $\eta^i = \rho\dot{x}^i$ is a solution of equa-

tions (24.1). Moreover, the functions $L_i(\eta)$ are not independent since they satisfy the conditions $\dot{x}^i L_i(\eta) \equiv 0$. These results are immediate consequences of the identities

$$\dot{x}^i f_{\dot{x}^i} = f, \qquad \dot{x}^i f_{\dot{x}^i\dot{x}^k} \equiv 0, \qquad \dot{x}^i f_{\dot{x}^i x^k} = f_{x^k},$$

which hold because of the homogeneity relations $f(x, k\dot{x}) = kf(x, \dot{x})$ $(k > 0)$.

In order to establish the analogue of Theorem 10.1, it is convenient to introduce first the notion of special accessory extremals. To do so, let $\varphi(x, \dot{x})$ be a function of class C^3 which is different from zero along E_{12}, is positively homogeneous of order one in the variables $(\dot{x})$, and is such that the equations

$$\varphi_{x^i} - (d/dt)\, \varphi_{\dot{x}^i} = 0$$

hold along E_{12}. Such a function can be constructed in many ways. For example, if $f(x, \dot{x})$ is different from zero along E_{12}, we may choose $\varphi = f$, or if $W(x)$ is a function of class C^4 such that $dW > 0$ along E_{12} we can choose $\varphi = W_{x^i}\, \dot{x}^i$, as one readily verifies. It should be noted that the determinant

$$\begin{vmatrix} f_{\dot{x}^i\dot{x}^k} & \varphi_{\dot{x}^i} \\ \varphi_{\dot{x}^k} & 0 \end{vmatrix} \tag{24.2}$$

is different from zero on E_{12} if and only if the matrix $\| f_{\dot{x}^i\dot{x}^k} \|$ has rank n on E_{12}. This fact can be established readily with the help of the following well-known consequences

$$\dot{x}^i f_{\dot{x}^i\dot{x}^k} \equiv 0, \qquad \dot{x}^i \varphi_{\dot{x}^i} = \varphi$$

of the positive homogeneity of the functions f and φ.

We now define a *special accessory extremal*[29] to be an accessory extremal which satisfies the equations

$$(d/dt)(\varphi_{\dot{x}^i}\eta^i) = \varphi_{x^i}\eta^i + \varphi_{\dot{x}^i}\dot{\eta}^i = a = \text{constant}. \tag{24.3}$$

It is easily seen by direct substitution that a special accessory extremal η^i is of the form $\eta^i = \rho x^i$ if and only if ρ is of the form $\rho = (at + b)/\varphi$, where a, b are constants.

A value $t_3 \neq t_1$ will be said to define a point 3 conjugate to the point 1 on E_{12} if there exists a special accessory extremal having $\eta^i(t_1) = \eta^i(t_3) = 0$ and $(\eta) \not\equiv (0)$ on t_1t_3. By the *order of* 3 *as a* conjugate point of 1 will be meant the number of linearly independent special accessory extremals in a maximal set having $\eta^i(t_1) = \eta^i(t_3) = 0$ and $(\eta) \not\equiv (0)$ on t_1t_3. It will be convenient to say that t_3 is conjugate to t_1 if t_3 defines a point 3 conjugate to 1 on E_{12}.

Every special accessory extremal satisfies with $\lambda = 0$ the equations

$$L_i(\eta) + \lambda\varphi_{\dot{x}^i} = 0, \qquad (d^2/dt^2)(\varphi_{\dot{x}^i}\eta^i) = 0. \tag{24.4}$$

[29] Bliss, *Jacobi's condition for problems of the Calculus of Variations in parametric form*, Trans. Amer. Math. Soc., vol. 17 (1916), pp. 195–206. Hestenes, *A note on the Jacobi condition for parametric problems in the Calculus of Variations*, Bull. Amer. Math. Soc., vol. 40 (1934), pp. 297–302.

Conversely, every solution η^i, λ of class C'' of these equations has $\lambda = 0$ since

$$\dot{x}^i L_i(\eta) \equiv 0\,, \qquad \dot{x}^i \varphi_{\dot{x}^i} = \varphi \neq 0\,,$$

and hence the functions η^i define a special solution. The equations (24.4) therefore completely characterize special accessory extremals.

When the determinant (24.2) is different from zero, the equations (24.4) can be solved for the variables $\ddot{\eta}^i$, λ. The solutions have the form $\lambda \equiv 0$ (since $\dot{x}^i L_i(\eta) \equiv 0$) and

$$\ddot{\eta}^i = A^i_k(t)\eta^k + B^i_k(t)\dot{\eta}^k\,.$$

It follows that special accessory extremals have properties like those of ordinary accessory extremals described in §1. For example, there are $2n + 2$ linearly independent special accessory extremals $\eta^{is}(t)$ $(s = 1, \cdots, 2n + 2)$ such that every special accessory extremal η^i is expressible linearly with constant coefficients in terms of these $2n + 2$ extremals. Moreover, the conjugate points of t_1 are determined by the zeros $t_3 \neq t_1$ of the determinant

$$\begin{vmatrix} \eta^{is}(t) \\ \eta^{is}(t_1) \end{vmatrix} \qquad (s = 1, \cdots, 2n + 2)\,.$$

It follows that if t_2 is not conjugate to t_1, the points (t_1, η_1) and (t_2, η_2) can be joined by a unique special accessory extremal. Similarly, there are $n + 1$ linearly independent accessory extremals $\eta^{ik}(t)$ $(k = 0, 1, \cdots, n)$ having $\eta^{ik}(t_1) = 0$ and such that every accessory extremal η^i having $\eta^i(t_1) = 0$ is expressible linearly in terms of these extremals and such that the conjugate points of t_1 are determined by the zeros $t_3 \neq t_1$ of the determinant $|\,\eta^{ik}(t)\,|$.

In this case the *condition of Legendre* will be said to hold on E_{12} if at each element $(x, \dot{x})$ on E_{12} the inequality

$$f_{\dot{x}^i\dot{x}^k}\,\pi^i\pi^k \geqq 0 \qquad (i, k = 0, 1, \cdots, n)$$

is true for every set of constants $(\pi) \neq (\rho\dot{x})$. If the equality sign can be excluded then the *strengthened condition of Legendre* is said to hold on E_{12}.

It was seen in §3 that if the type number of $J_2(\eta)$ is finite, the condition of Legendre must hold. Conversely, if we set

$$I_2(\eta) = \int_{t_1}^{t_2} 2\omega(\eta, \dot{\eta})\,dt\,,$$

$$I_2(\xi, \eta) = \int_{t_1}^{t_2} \Omega(\xi, \eta)\,dt\,,$$

We have the following result:

THEOREM 24.1. *Let s be the sum of the orders of the conjugate points of t_1 between t_1 and t_2. If the strengthened condition of Legendre holds, there exists a set of s (and no fewer) natural isoperimetric conditions*

$$I_2(\xi_\alpha, \eta) = 0 \qquad (\alpha = 1, \cdots, s)$$

such that $I_2(\eta) \geqq 0$ *for every admissible arc* $\eta^i(t)$ *vanishing at* t_1 *and* t_2 *and satisfying these s conditions. The equality holds only in case* η^i *is expressible in the form* $\eta^i = u^i - \rho x^i$ *where* u^i *is a special accessory extremal having* $u^i(t_1) = u^i(t_2) = 0$ *and* $\rho(t)$ *is a function of class* D' *vanishing at* t_1 *and* t_2.

The proof of this result is like that of Theorem 5.1. Lemma 4.1 is true here also if the conjugate system η_{ik} there used is replaced by a set of linearly independent special accessory extremals η^{ik} having $\eta^{ik}(t_1) = 0$, as one readily verifies. The proof of the fact will be simplified if one notes that one of the extremals $\eta^{ik}(t)$, say the first, can be chosen to be the special accessory extremal $\eta^{i0} = (t - t_1)\dot{x}^i/\varphi$, and that the remaining n solutions can be chosen so as to satisfy the condition $\varphi_{\dot{x}^i}\eta^i = 0$. Moreover, in this case the equation (4.4) takes the form

$$I_2(\eta) = \int_{t_1}^{t_2} f_{\dot{x}^i\dot{x}^k}\, v^i v^k \, dt \geqq 0,$$

where $v^i = \eta^{ik}a_k'$. The equality holds only in case $v^i = \sigma(t)\dot{x}^i$. From the relation

$$\varphi_{\dot{x}^i} v^i = \sigma(t)\varphi$$

we conclude that $(t - t_1)a_0' = \sigma(t)\varphi$ and hence that the derivatives $a_1', \cdots, a_n'$ are zero. It follows that in this case the functions $a_1, \cdots, a_n$ are all constants and $a_0 = \rho(t)$ vanishes at t_1 and t_2 and hence that η^i is expressible in the form described in the last statement of the theorem. The remainder of the proof can now be made by the arguments like those used in the proof of Theorem 5.1.

We now return to the functional $J_2(\eta)$ and define the *order of concavity* of $J_2(\eta)$ formally as in §7 using special accessory extremals instead of the accessory extremals there used. With these definitions in mind we can prove the following:

THEOREM 24.2. *Let r be the order of concavity of* $J_2(\eta)$ *and s the sum of the orders of the conjugate points of* t_1 *between* t_1 *and* t_2. *If the strengthened condition of Legendre holds, there exists a set of* $m = r + s$ *natural isoperimetric conditions*

$$J_2(\xi_\alpha, \eta) = 0 \qquad (\alpha = 1, \cdots, m) \tag{24.5}$$

such that the inequality $J_2(\eta) \geqq 0$ *holds for every set of admissible variations* η^i *satisfying the m conditions, the equality holding only in case* η^i *is expressible in the form* $\eta^i = u^i - \rho x^i$, *where* u^i *is a special accessory extremal satisfying the conditions* $\psi_\mu = 0$ *and* $\rho(t)$ *is a function of class* D' *having* $\rho(t_1) = \rho(t_2) = 0$. *Moreover, the inequality*

$$J_2(\xi_\alpha, \xi_\beta)a_\alpha a_\beta < 0 \qquad (\alpha, \beta = 1, \cdots, m)$$

is true for every set of constants $(a) \neq (0)$.

The proof is like that of Theorem 10.1 with the help of Theorem 24.1.

COROLLARY 1. *The set* (24.5) *forms a minimal set for the functional* $J_2(\eta)$.

25. **Natural isoperimetric conditions.** In this section we assume that E_{12} is an extremal arc (not intersecting itself) satisfying the end conditions $\psi_\mu = 0$ and the transversality conditions (23.3).

Let $\xi^i(x)$ be a contravariant vector of class C^4 such that[30]

$$\Psi_\mu(\xi) = \psi_{\mu x_1^i}\xi^i(x_1) + \psi_{\mu x_2^i}\xi^i(x_2) = 0$$

on the manifold $\psi_\mu = 0$. The condition

$$J_1(\xi) = g_{x_1^i}\xi^i(x_1) + g_{x_2^i}\xi^i(x_2) + \int_{t_1}^{t_2} \{f_{x^i}\xi^i + f_{\dot{x}^i}\dot{\xi}^i\}\, dt = 0\,,$$

where $\dot{\xi}^i = \xi^i_{x^k}\dot{x}^k$, will be called a *natural isoperimetric condition.* It is satisfied by every extremal arc which satisfies the conditions $\psi_\mu = 0$ and the transversality conditions (23.3), as one readily verifies.

Consider now a set of m natural isoperimetric conditions

$$J_1(\xi_\alpha) = 0 \qquad (\alpha = 1, \cdots, m). \tag{25.1}$$

Such a set will be called a *minimal set* for the extremal arc E_{12} if E_{12} affords at least a weak minimum in the class of admissible arcs satisfying these m conditions (25.1), and if no proper subset of these conditions has this property. The set (25.1) will be called a *proper set* for E_{12} if the determinant

$$|\, J_2(\xi_\alpha, \xi_\beta)\, |$$

is different from zero on E_{12}. In this case there exist infinitely many admissible arcs satisfying the conditions (24.1) in every neighborhood of E_{12}, as follows from

LEMMA 25.1. *If the set* (25.1) *forms a proper minimal set for E_{12}, for every admissible variation $\eta^i(t)$ satisfying the conditions*

$$J_2(\xi_\alpha, \eta) = 0 \qquad (\alpha = 1, \cdots, m)$$

there exists a one-parameter family of admissible arcs

$$x^i = x^i(t, b)$$

satisfying the conditions (25.1), *containing E_{12} for $b = 0$ and having η^i as its variations along E_{12}. The derivatives x^i_b, x^i_{tb}, x^i_{bb}, x^i_{tbb} exist and are continuous except possibly at a finite number of points on t_1t_2.*

The proof is like that of Lemma 19.1. Let $\eta^i_\gamma(t)$ $(\gamma = 1, \cdots, m + p)$ be a set of $m + p$ admissible variations whose determinant

$$\begin{vmatrix} J_2(\xi_\alpha, \eta_\gamma) \\ \Psi_\mu(\eta_\gamma) \end{vmatrix} \tag{25.2}$$

[30] In order to discuss the general case completely one must choose the functions ξ^i of the form $\xi^i(x, x_1, x_2)$. However, the functions $\xi^i(x)$ are sufficiently general for the separated end point case. We shall accordingly restrict ourselves to this simple case.

is different from zero on E_{12}. We substitute the functions

$$x^i(t) + a_\gamma \eta^i_\gamma(t) + b\eta^i(t)$$

in equations (25.1) and $\psi_\mu = 0$ and obtain a set of functions $J_{1\alpha}(a, b)$, $\psi_\mu(a, b)$ having $J_{1\alpha}(0, 0) = 0$. The functional determinant of $J_{1\alpha}(a, b)$, $\psi_\mu(a, b)$ with respect to the variables $a_1, \cdots, a_{m+p}$ at $(a, b) = (0, 0)$ is precisely the determinant (25.2) and is accordingly different from zero. The equations $J_{1\alpha}(a, b) = 0$, $\psi_\mu(a, b) = 0$ have solution $a_\gamma = A_\gamma(b)$ of class C'', with $A_\gamma(0) = 0$. Moreover, if we differentiate the expression $J_{1\alpha}[A(b), b]$, we find, as in the proof of Lemma 19.1, that $A'_\gamma(0) = 0$. It follows that the family

$$x^i = x^i(t) + A_\gamma(b)\eta^i_\gamma(t) + b\eta^i(t)$$

has the properties described in the lemma, as one readily verifies.

The arc E_{12} will be said to satisfy the *condition of Weierstrass* if at each element $(x, \dot{x})$ on E_{12} the inequality

$$E(x, \dot{x}, \dot{y}) \geqq 0$$

holds for every non-null set $(\dot{y}) \neq (kx)$ $(k > 0)$, where

$$E(x, \dot{x}, \dot{y}) = f(x, \dot{y}) - \dot{y}^i f_{\dot{x}^i}(x, \dot{x}). \tag{25.3}$$

We have further

THEOREM 25.1. *If the set* (25.1) *forms a proper minimal set for the extremal E_{12} and E_{12} affords a strong minimum relative to neighboring admissible arcs satisfying these conditions, then E_{12} satisfies the condition of Weierstrass.*

This result follows at once from Theorem 22.1. For, if we consider our problem as a non-parametric problem in (tx)-space, the image of E_{12} must be a minimizing arc for this problem and hence we have

$$f(x, \dot{y}) - f(x, \dot{x}) - (\dot{y}^i - \dot{x}^i) f_{\dot{x}^i}(x, \dot{x}) \geqq 0.$$

But this function is precisely the E-function (25.3) since $f = \dot{x}^i f_{\dot{x}^i}$.

The condition of Legendre has been defined in the last section. We now have

THEOREM 25.2. *If the set* (25.1) *forms a proper minimal set for the extremal arc E_{12}, then the condition of Legendre holds on E_{12}. If E_{12} is non-singular, the strengthened condition of Legendre holds on E_{12}.*

This result follows from Theorem 25.1 in the same manner as Theorem 19.3 follows from Theorem 19.2.

THEOREM 25.3. *If the set* (25.1) *forms a proper minimal set for E_{12}, the inequality $J_2(\eta) \geqq 0$ holds for every set of admissible variations η^i satisfying the conditions*

$$J_2(\xi_\alpha, \eta) = 0 \qquad (\alpha = 1, \cdots, m).$$

The proof is like that of Theorem 19.4 with the help of Lemma 25.1.

26. A fundamental theorem. In this section we shall establish an extended sufficiency theorem of apparently new type for the fixed end point case. This theorem will be useful in our final sufficiency theorems to be given in the next section.

We begin with the notion of Mayer field. A *Mayer field* is a region $\mathfrak{F}$ on our Riemannian manifold $\mathfrak{R}$ together with a set of contravariant vectors $p^i(x)$ with $(p) \neq (0)$ and of class C' on $\mathfrak{F}$ such that the Hilbert integral

$$I^* = \int f_{\dot{x}^i}[x, p(x)]\, dx^i$$

is independent of the path in $\mathfrak{F}$. The solutions of class C'' of the equations $\dot{x}^i = p^i(x)$ are called *extremals of the field.* Along such an extremal E we have $I^*(E) = I(E)$, where

$$I = \int_{t_1}^{t_2} f(x, \dot{x})\, dt\,.$$

The following basic lemma is well-known and follows at once from the independence of the path of the Hilbert integral I^*. The function E used in this lemma is, of course, the Weierstrass E-function (25.3).

LEMMA 26.1. *If E is an extremal of a field $\mathfrak{F}$ at each point of which the inequality*

$$E[x, p(x), \dot{x}] > 0 \tag{26.1}$$

holds for every admissible set $(x, \dot{x}) \neq (x, kp)$ $(k > 0)$, the inequality $I(C) > I(E)$ holds for every admissible arc C in $\mathfrak{F}$ joining the end points of E but not identical with E.

An extremal E_{12} will be said to satisfy the *strengthened condition of Weierstrass* if at each element $(x, \dot{x})$ on E_{12}, the inequality

$$E(x, \dot{x}, \dot{y}) > 0 \tag{26.2}$$

holds for every non-null set $(\dot{y}) \neq (k\dot{x})$ $(k > 0)$, where E is the Weierstrass E-function. The strengthened condition of Legendre has been described in §24.

LEMMA 26.2. *If the extremal E_{12} satisfies the strengthened condition of Weierstrass and Legendre, the inequality* (26.2) *holds for every set $(x, \dot{x})$ in a neighborhood N of those on E_{12} and for all values of $(\dot{y}) \neq (k\dot{x})$ $(k > 0)$.*

The proof of this lemma is well-known.

We have the following analogue of Theorem 20.1:

LEMMA 26.3. *If an extremal arc E_{12} satisfies the strengthened conditions of Weierstrass and Legendre and has on it no point conjugate to its initial end point* 1, *there exists a neighborhood $\mathfrak{F}$ of E_{12} and neighborhoods N_1, N_2 of the end points* 1 *and* 2, *respectively, such that every extremal E in $\mathfrak{F}$ with ends in N_1, N_2 affords a proper minimum of the functional I in the class of admissible arcs in $\mathfrak{F}$ joining its end points. Moreover, every sub-arc of E affords a proper minimum relative to admissible arcs in $\mathfrak{F}$ joining its end points.*

For suppose the coördinate system (x) has been chosen so that $\dot{x}^0(t) > 0$

along E_{12}. We may accordingly choose $t = x^0$ as the parameter for E_{12} and and also for the neighboring extremals. These extremals will accordingly be the extremals for the non-parametric integral

$$I = \int_{x_1}^{x_2} F(x, y, y')\, dx ,$$

where $(x^0, x^1, \cdots, x^n) = (x, y_1, \cdots, y_n)$ and

$$F(x, y, y') = f(x^0, \cdots, x^n, 1, \dot{x}^1, \cdots, \dot{x}^n) .$$

The Legendre condition for the non-parametric integral is clearly the same as that for the parametric integral. Similarly, the conjugate points of 1 are the same for the two problems. Hence it follows from the proof of Lemma 20.1 that there exist neighborhoods $\mathfrak{F}$ of E_{12} and N_1, N_2 of the ends of E_{12} such that every extremal in $\mathfrak{F}$ with ends in N_1, N_2 is an extremal of a Mayer field defined over $\mathfrak{F}$. Moreover, if $\mathfrak{F}$ is taken sufficiently small the condition (26.1) will hold in each of these fields, by virtue of Lemma 26.2. The lemma now follows from Lemma 26.1.

We come now to our fundamental theorem. We assume, of course, that E_{12} does not intersect itself.

THEOREM 26.1. *Let E_{12} be an extremal arc satisfying the strengthened conditions of Weierstrass and Legendre and let E_{34} be a sub-arc of E_{12} having on it no point conjugate to its initial point* 3. *There exists a neighborhood $\mathfrak{F}$ of E_{12} such that E_{34} affords a proper minimum of the integral I relative to admissible arcs in $\mathfrak{F}$ joining its end points.*

In order to obtain this result, we choose a constant d so small that no sub-arc of E_{12} whose length does not exceed d has on it a pair of conjugate points. This can be done by virtue of the fact that the conjugate points of a given point are isolated.

Let $P_0, P_1, \cdots, P_{q+1}$ be a set of successive points on E_{12} such that the length of the sub-arc P_jP_{j+3} does not exceed d. We may suppose the point P_0 lies a little to the left of the point 1 and P_{q+1} lies a little to the right of the point 2 and that for some value of k the points P_k, P_{k+l} coincide with the points 3 and 4 of the sub-arc E_{34} in question. Through the points P_j pass analytic manifolds π_j cutting E_{12} orthogonally. From Lemma 26.3 it follows readily that there exists a neighborhood $\mathfrak{F}'$ of E_{12} such that not only E_{34} but also every extremal E in $\mathfrak{F}'$ with ends on the manifolds P_{j-1}, P_{j+2} affords a proper minimum of I relative to admissible arcs in $\mathfrak{F}'$ joining its end points but not crossing the manifolds passing through its end points. Let $\mathfrak{F}$ be a neighborhood of E_{12} interior to $\mathfrak{F}'$ such that every pair of points Q_j, Q_{j+1} in $\mathfrak{F}$ on the manifolds π_j, π_{j+1}, respectively, determine an extremal segment E in $\mathfrak{F}'$ with ends on the hyperplanes π_{j-1} and π_{j+2}. Let E_j be the segment of E between Q_j and Q_{j+1}. It is clear that E_j affords a proper minimum to the integral I relative to admissible arcs in $\mathfrak{F}'$ joining its end points but not crossing the manifolds π_{j-1}, π_{j+2}.

With the help of these preliminary constructions we can prove the theorem as follows. Let C be an admissible arc in $\mathfrak{F}$ joining the end points of the sub-arc E_{34}. It is clear that the curve C crosses the manifolds π_j at most a finite number of times. If C does not cross the manifolds π_k and π_{k+l}, which pass through the end points of E_{34}, then $I(C) > I(E_{34})$, unless $C \equiv E_{34}$, by virtue of the remarks made in the last paragraph. Suppose now that C crosses the hyperplane π_k. Let π_{h-1} be the first manifold on the left which C does not cross. As the point P moves along the arc C from the point 3 to 4 it will cross the manifold π_h at a first point Q_h and will subsequently reach the manifold π_{h+1} at a first point Q_{h+1}. Let C' be the segment of C bounded by the Q_h and Q_{h+1} and let E be the extremal joining its end points. The arc C' is not identical with E, since otherwise C could not cross π_h at the point Q_h. Hence we have $I(C') > I(E)$ according to the remarks made in the last paragraph. If now we replace the sub-arc C' of C by E, we obtain a new curve C_1 with $I(C_1) < I(C)$. The segment E of C_1 may not lie wholly in $\mathfrak{F}$, but this fact leads to no difficulties since the arc C_1 crosses the manifolds π_j at points in $\mathfrak{F}$. It is clear that the arc C_1 crosses π_h at most $r - 2$ times, where r denotes the number of times the arc C crosses π_h. By a repetition of this process to the arc C_1, and so on, we finally obtain a curve C_h which does not cross π_h. Proceeding in this manner we obtain next a curve C_{h+1} which does not cross the manifold π_{h+1} and finally a curve C_2 which does not cross the manifold π_k passing through the point 3 on E_{34}. Clearly $I(C_2) < I(C)$. In similar manner we may replace the curve C_2 by a curve C_3 in $\mathfrak{F}'$ which joins the end points of E_{34} but does not cross either of the manifolds π_k or π_{k+l} passing through the end points of E_{34}. We have accordingly

$$I(C) > I(C_3) \geqq I(E_{12})$$

and the theorem is established.

Corollary. *Every extremal sub-arc in $\mathfrak{F}'$ joining points in $\mathfrak{F}$ lying on successive manifolds π_j, π_{j+1} affords a proper minimum of J relative to admissible arcs in $\mathfrak{F}$ joining its end points.*

The results described in Theorem 26.1 can be extended readily to the case in which E_{34} has pairs of conjugate points on it provided that its end points are not conjugate. This can be done by adjoining a suitably chosen set of natural isoperimetric conditions for the arc E_{34}. The key point in the proof lies in the fact that we can choose these natural isoperimetric conditions to be identically zero on the left of the manifold π_k passing through the point 3 and on the right of the manifold π_{k+l} passing through the point 4.

27. The existence of minimal sets. Consider now a set of natural isoperimetric conditions

$$(27.1) \qquad J_1(\xi_\alpha) = g_\alpha(x_1, x_2) + \int_{t_1}^{t_2} f_\alpha(x, \dot{x})\, dt = 0 \qquad (\alpha = 1, \cdots, m)\,.$$

Let $F = f + \lambda_\alpha f_\alpha$ and $E_\lambda(x, \dot{x}, \dot{y})$ be the Weierstrass E-function formed for the function F.

We shall need the following lemma which can be established by the usual arguments:

LEMMA 27.1. *If an extremal arc E_{12} satisfies the strengthened condition of Weierstrass and Legendre, there exists a neighborhood N of the values $(x, \dot{x}, \lambda)$ on E_{12} such that the inequality*

$$E_\lambda(x, \dot{x}, \dot{y}) > 0$$

is true for all values $(x, \dot{x}, \lambda)$ in N and every non-null set $(\dot{y}) \neq (k\dot{x})$ $(k > 0)$.

Let

$$I_\lambda = \int_{t_1}^{t_2} (f + \lambda_\alpha f_\alpha)\, dt$$

and suppose that the arc E_{12} satisfies the strengthened condition of Weierstrass and Legendre. With the help of the last lemma it is seen that the analogues of Lemmas 20.1 and 26.3 are true also for the functional I_λ with (λ) in a neighborhood Λ of $(\lambda) = (0)$. Suppose now that E_{12} be cut orthogonally by regular analytic manifolds $\pi_0, \pi_1, \cdots, \pi_{q+1}$ as described in the proof of Theorem 26.1. By an argument like that used in the proof of Theorem 26.1 and its corollary one obtains the following

LEMMA 27.2. *Under the above hypotheses there exist neighborhoods $\mathfrak{F}'$ of E_{12} and Λ of $(\lambda) = (0)$ such that every isoperimetric extremal arc E_λ in $\mathfrak{F}'$ with multipliers λ_α in Λ and joining a pair of points on successive manifolds π_k and π_{k+1} affords a proper minimum to I_λ relative to admissible arcs in $\mathfrak{F}'$ joining its end points.*

By the order of concavity of E_{12} will be meant the order of concavity of the second variation $J_2(\eta)$ along E_{12}. The arc E_{12} will be said to be non-degenerate if the order of degeneracy of $J_2(\eta)$ along E_{12} is zero.

In the following theorem we assume that the end conditions $\psi_\mu = 0$ are of the form

$$\psi_{\mu_1}[x(t_1)] = 0, \quad \psi_{\mu_2}[x(t_2) = 0 \quad (\mu_1 = 1, \cdots, p_1 \leqq n; \quad \mu_2 = 1, \cdots, p_2 \leqq n).$$

We assume further that E_{12} is not tangent to these end manifolds and that the end points of E_{12} are the only pairs of points on E_{12} satisfying the conditions $\psi_\mu = 0$.

THEOREM 27.1. *Let E_{12} be a non-degenerate extremal arc (not intersecting itself) satisfying the end conditions $\psi_\mu = 0$ and the transversality conditions* (23.3). *Let r be the order of concavity and s the sum of the orders of the conjugate points of the point* 1 *between* 1 *and* 2. *If E_{12} satisfies the strengthened conditions of Weierstrass and Legendre, there exists a set of $m = r + s$ (and no fewer) natural isoperimetric conditions* (27.1) *such that E_{12} affords a proper strong minimum to J relative to neighboring admissible arcs satisfying the conditions $\psi_\mu = 0$ and* (27.1).

In order to obtain this result, we suppose that the arc E_{12} is cut orthogonally

in points $P_1, \cdots, P_q$ by regular analytic manifolds $\pi_1, \cdots, \pi_q$, as described in the proof of Theorem 26.1, replacing the manifolds π_0, π_{q+1} by $\psi_{\mu_1} = 0, \psi_{\mu_2} = 0$, respectively. We shall denote the value of the parameter t of E_{12} at the point P_h by t_h and set $\tau_0 = t_1, \tau_{q+1} = t_2$. We next select a minimal set

$$J_2(\xi_\alpha, \eta) = 0 \qquad (\alpha = 1, \cdots, m)$$

for the second variation $J_2(\eta)$ along E_{12}. It is clear from Theorem 24.2 that the number of the conditions in this set is precisely equal to the desired sum $r + s$. We may suppose that the functions $\xi_\alpha^i(t)$ are of class C^4.

Let $\eta_\alpha^i(t)$ be a set of broken accessory extremals having its corners at $t = \tau_h$ such that

$$\eta_\alpha^i(\tau_h) = \xi_\alpha^i(\tau_h) \qquad (h = 0, 1, \cdots, q + 1)\,. \tag{27.2}$$

If necessary, we can modify the functions $\xi_\alpha^i(t)$ so that the inequality

$$J_2(\eta_\alpha, \eta_\beta)a_\alpha a_\beta < J_2(\xi_\alpha, \xi_\beta)a_\alpha a_\beta < 0$$

holds for every set of constants $(a) \neq (0)$. It follows that the determinant

$$|\, J_2(\eta_\alpha, \eta_\beta) - J_2(\xi_\alpha, \xi_\beta) \,| \tag{27.3}$$

is different from zero on E_{12}.

Let $\xi_\alpha^i(x)$ be a set of $m = r + s$ functions of class C^4 on $\mathfrak{R}$ satisfying the conditions $\Psi_\mu(\xi_\alpha) = 0$ on the manifold $\psi_\mu = 0$ and having $\xi_\alpha^i[x(t)] = \xi_\alpha^i(t)$ along E_{12}. The isoperimetric conditions obtained by substituting these functions in equations (27.1) can now be shown to have the properties described in the theorem. In order to do so, let the equations of the manifolds π_h be given by

$$x^{ih} = X^{ih}(b_{1h}, \cdots, b_{nh}) \qquad (h = 1, \cdots, q)\,, \tag{27.4}$$

the parameters $(b_{ih}) = (0)$ defining the point P_h on E_{12}. Since the manifolds are regular, the determinant $|\, \partial X^{ih}/\partial b_{jh} \,|$ is different from zero for $(b) = (0)$ and for all h. It is clear also that end conditions $\psi_\mu = 0$ are equivalent to conditions of the form

$$x^{is} = x^{is}(b_1, \cdots, b_\sigma) \qquad (s = 1, 2) \tag{27.5}$$

with $|\, x_{b_\nu}^{is}(0) \,| \neq 0$ the end points of E_{12} being given by the values $(b_\nu) = (0)$. Let

$$x^i = x^i(t, b_1, \cdots, b_\sigma, b_{11}, \cdots, b_{qn}, \lambda_1, \cdots, \lambda_m) \qquad (t_1 \leqq t \leqq t_2) \tag{27.6}$$

be a $(\sigma + qn + m)$-parameter family of broken isoperimetric extremals for the integral

$$I_\lambda = \int_{t_1}^{t_2} F\, dx = \int_{t_1}^{t_2} (f + \lambda_\alpha f_\alpha)\, dt$$

having its corners on the manifolds π_h, satisfying the conditions $\psi_\mu = 0$, and containing E_{12} for $(b, \lambda) = (0, 0)$. We may suppose that this family has been

chosen so that

$$x^i(t_s, b, \lambda) = x^{is}(b), \qquad x^i(\tau_h, b, \lambda) = X^{ih}(b) \qquad (s = 1, 2; h = 1, \cdots, q).$$

It follows that if we set $\bar{\eta}^i_\alpha = (\partial/\partial\lambda_\alpha)x^i(t, 0, 0)$, we have

$$\bar{\eta}^i_\alpha(\tau_h) = 0 \qquad (h = 0, 1, \cdots, q + 1). \tag{27.7}$$

Moreover, if we substitute the functions $x^i(t, b, \lambda)$ in the Euler equations

$$F_{x^i} - (d/dt)F_{\dot{x}^i} = P_i + \lambda_\alpha \xi^k_{\alpha x^i} P_k + \lambda_\alpha L_i(\xi_\alpha) = 0$$

for I_λ, where P_i, L_i are defined by equations (23.1), (24.1), and differentiate with respect to λ_α, it is found that for $(b, \lambda) = (0, 0)$ we have

$$L_i(\bar{\eta}_\alpha + \xi_\alpha) = 0$$

along E_{12}. The arc $\eta^i_\alpha = \bar{\eta}^i_\alpha + \xi^i_\alpha(t)$ is therefore a broken accessory extremal satisfying the conditions (27.2) by virtue of equations (27.7). It follows that the functions differ from those described in the last paragraph by at most $\rho\dot{x}^i$ with $\rho(t_h) = 0$ $(h = 0, 1, \cdots, q + 1)$ and hence can be taken as identical. By the use of equations (27.2) and the usual integration by parts it is found that

$$J_2(\xi_\alpha, \eta_\beta) = J_2(\eta_\alpha, \eta_\beta)$$

and hence that

$$\begin{aligned} J_2(\xi_\alpha, \bar{\eta}_\beta) &= J_2(\xi_\alpha, \eta_\beta) - J_2(\xi_\alpha, \xi_\beta) \\ &= J_2(\eta_\alpha, \eta_\beta) - J_2(\xi_\alpha, \xi_\beta). \end{aligned}$$

The determinant

$$| J_2(\xi_\alpha, \bar{\eta}_\beta) | \tag{27.8}$$

is therefore equal to the determinant (27.4) and hence must be different from zero.

When the family (27.6) is substituted in the functions $J_1(\xi_\alpha)$ a set of m functions $J_{1\alpha}(b, \lambda)$ is obtained having $J_{1\alpha}(0, 0) = 0$. Moreover, the functional determinant of $J_{1\alpha}(b, \lambda)$ with respect to $\lambda_1, \cdots, \lambda_m$ at $(b, \lambda) = (0, 0)$ is precisely the determinant (27.8), and is accordingly different from zero. The equations $J_{1\alpha}(b, \lambda) = 0$ therefore have solutions $\lambda_\alpha = \lambda_\alpha(b)$ of class C'' with $\lambda_\alpha(0) = 0$. The family

$$x^i = x^i(t, b) = x^i[t, b, \lambda(b)], \qquad \lambda_\alpha = \lambda_\alpha(b) \qquad (t_1 \leqq t \leqq t_2), \tag{27.9}$$

of broken isoperimetric extremals satisfies the equations (27.1) and $\psi_\mu = 0$ and contains E_{12} for $(b) = (0)$. Renumber the b's, $b_1, \cdots, b_{qn+m}$. The variations

$$\eta^i = \delta x^i = x^i{}_{b_j}(t, 0)db_j \qquad (j = 1, \cdots, qn + m)$$

accordingly satisfy the conditions $J_2(\xi_\alpha, \eta) = 0$ and $\Psi_\mu(\eta) = 0$. Hence $J_2(\delta x) \geqq 0$. Since E_{12} is non-degenerate, the equality holds only in case δx^i is of the form $\rho\dot{x}^i$, where $\rho(t_1) = \rho(t_2) = 0$. But this is impossible unless $\rho \equiv 0$

by virtue of the fact that the arc E_{12} is not tangent to the manifolds π_h. It follows that when the functions $x^i(t, b)$ are substituted in the functional J the function $J(b)$ so obtained satisfies the condition $J(b) > J(0)$ for $(b) \neq (0)$ and sufficiently near $(b) = (0)$ since $dJ = 0$, $d^2J = J_2(\delta x) > 0$ at $(b) = (0)$ for $(db) \neq (0)$.

We are now in position to complete the proof of Theorem 27.1. Let $\mathfrak{F}'$ be a neighborhood of E_{12} and Λ a neighborhood of $(\lambda) = (0)$ related to E_{12} as in Lemma 27.2. Let $\mathfrak{F}$ be a neighborhood of E_{12} interior to $\mathfrak{F}'$ such that every set of $q + 1$ points in $\mathfrak{F}$, one on each of the manifolds (27.4), (27.5), determines a set of parameters (b) such that $J(b) > J(0)$, $(b) \neq (0)$, and such that the broken extremal of the family (27.9) determined by (b) is in $\mathfrak{F}'$ and $\lambda_\alpha(b)$ is in Λ. Consider now an admissible C in $\mathfrak{F}$ and let Q_h be one of the points in which the curve C cuts the manifold π_h. The points Q_h and the end values of E_{12} determine a set of parameters (b), which in turn determines a broken isoperimetric extremal E_b of the family (27.9). Since the arc E_b passes through the points Q_h on C and joins the end points of the arc C, we have

$$I_\lambda(C) > I_\lambda(E_b),$$

unless $C \equiv E_b$, the multipliers λ_α being those belonging to E_b. Since the arcs C and E_b satisfy the conditions (27.1) and have the same end values, it follows from the last relation and the above remarks that

$$J(C) \geqq J(E_b) = J(b) \geqq J(0) = J(E_{12}),$$

the equality holding only in case $C \equiv E_{12}$, as was to be proved.

It is clear that none of the conditions (27.1) can be dropped, since otherwise $J_2(\xi_\alpha, \eta) = 0$ could not be a minimal set for $J_2(\eta)$ along E_{12}. The set (27.1) accordingly forms a minimal set. From the above proof it follows readily that the set (27.1) forms a minimal set for E_{12} if and only if the equations of variations $J_2(\xi_\alpha, \eta) = 0$ form a minimal set for $J_2(\eta)$. Hence we have the following corollary.

COROLLARY. *If a non-degenerate extremal arc E_{12} satisfies the strengthened conditions of Legendre, then the number of conditions in a minimal set is the same for every such set.*

28. **More general natural isoperimetric conditions.** In the fixed end point case and in the case of closed extremals it is possible to define natural isoperimetric conditions of a very general type. In the fixed end point case we select a function $H(x, \dot{x})$ of class C^4 such that

$$(28.1) \qquad H(x, k\dot{x}) = H(x, \dot{x}) \ (k > 0), \qquad H(x_1, \dot{x}) = H(x_2, \dot{y})$$

for all non-null values of $(\dot{x})$ and $(\dot{y})$, where (x_1) and (x_2) are the coördinates of the fixed points 1 and 2. Let $\xi^i(x, \dot{x})$ be a solution of the equations

$$H_{\dot{x}^k} = \xi^i f_{\dot{x}^i \dot{x}^k}.$$

The condition

$$K(\xi) = \int_{t_1}^{t_2} \{f_{x^i}\xi^i + (H_{x^k} - \xi^i f_{\dot{x}^i x^k})\dot{x}^k\}\, dt = 0 \tag{28.2}$$

is a *general natural isoperimetric condition.* It is clear that this condition is satisfied by every extremal joining the points 1 and 2, since along such an arc we have

$$K(\xi) = \int_{t_1}^{t_2} \{dH/dt + \xi^i[f_{x^i} - (d/dt)f_{\dot{x}^i}]\}\, dt = 0\,.$$

If $f(x, \dot{x}) > 0$ we can always choose the functions ξ^i so that $H = \xi^i(x, \dot{x})f_{\dot{x}^i}$, at least in the non-singular case. The condition (28.2) can then be written in the simpler form

$$J_1(\xi) = \int_{t_1}^{t_2} \{f_{x^i}\xi^i + f_{\dot{x}^i}\xi^i_{x^k}\dot{x}^k\}\, dt = 0\,,$$

as one readily verifies. The natural isoperimetric conditions described in §25 are of the special type in which the function H is of the form $H = \xi^i(x)f_{\dot{x}^i}$.

It is clear that by the use of the general isoperimetric condition (28.2) minimal sets can be defined, as above, for a non-singular extremal arc. Moreover, the equation of variation of the equation (28.2) is expressible along E_{12} in the form

$$J_2(\bar{\xi}, \eta) = 0\,,$$

where $J_2(\xi, \eta)$ is the usual bilinear form for the second variation $J_2(\eta)$, and $\bar{\xi}^i = \xi^i[x(t), \dot{x}(t)]$. It is understood, of course, that $\eta^i(t_1) = \eta^i(t_2) = 0$. From this fact it follows readily that the theorems described in §§25 and 27 are still true in the fixed end point case if we replace the natural isoperimetric conditions there used by the more general conditions there defined. We have accordingly the following result.

THEOREM 28.1. *If the strengthened condition of Legendre holds on a non-degenerate extremal, the number of conditions in a minimal set of general isoperimetric conditions is always the same, and is equal to the sum of the orders of the conjugate points of the point* 1 *on* E_{12}.

In the case of closed extremals general isoperimetric conditions can be defined as above by omitting the second condition (28.1).

In the study of geodesics on a closed convex surface H. Poincaré showed that there is a shortest closed curve C on S which satisfies the condition

$$\iint K d\sigma = 2\pi \tag{28.3}$$

and that C is a geodesic. Here K denotes the Gaussian curvature of S and the integral is taken over one of the regions bounded by C. This condition can be transformed into one of the type here described. In fact, if we use isothermal

coördinates (x, y) (x being an angle of period 2π) so that the element of arc becomes

$$ds^2 = \lambda^2(dx^2 + dy^2),$$

condition (28.3) can be shown to be equivalent to a condition of the form

$$\int_C \{(\lambda_y x' - \lambda_x y')/\lambda - x'\}\, dt = 0.$$

This is a natural isoperimetric condition defined by the function

$$H = \operatorname{arc\,tan}\,(y'/x') - x.$$

V

The case of multiple integrals

Many of the results described in the preceding pages can be extended at once to multiple integrals. A complete theory for multiple integrals is not to be expected since such a theory does not exist even for the classical minimum. However, by the use of natural isoperimetric conditions one can go a long way towards a complete classification of extremals which do not satisfy the Jacobi condition. To illustrate this theory we shall use the non-parametric double integral problem.[31]

29. **The double integral problem.** Let A be an open region in the xy-plane bounded by a simply closed continuous arc C. Let S be a surface defined over $A + C$ by an equation of the form

$$z = z(x, y) \qquad [(x, y) \text{ on } A + C]\,. \tag{29.1}$$

In the following pages we shall be concerned with the classification of the extremals of an integral of the form

$$J = \iint_A f(x, y, z, p, q)\, dx\, dy,$$

where $p = \partial z/\partial x$, $q = \partial z/\partial y$.

The function $f(x, y, z, p, q)$ is assumed to be of class C^4 in a region $\mathfrak{R}$ of points (x, y, z, p, q). We shall suppose that the boundary C of A is composed of a finite number of regular sub-arcs. A surface S will be called a *regular surface* if the arc C bounding A in the xy-plane is regular and the function $z(x, y)$ is of class C^1, or if it is composed of a finite number of partial surfaces having these properties.

By an *admissible surface* will be meant a regular surface, all of whose elements (x, y, z, p, q) are in $\mathfrak{R}$, and which passes through a fixed curve L. The curve L

[31] In this part free use has been made of the unpublished lectures on multiple integrals given by Bliss at the University of Chicago, summer, 1933.

is assumed to be a simply closed continuous curve composed of a finite number of regular sub-arcs, and having the boundary C of A as its projection in the xy-plane.

The Euler equations of the integral J are the equations

$$f_z - (\partial/\partial x)f_p - (\partial/\partial y)f_q = 0 .$$

A solution (29.1) of these equations of class C'' with elements (x, y, z, p, q) in $\mathfrak{R}$ will be called an *extremal surface.*

30. **The accessory problem.** Along an extremal surface S the second variation of J is expressible in the form

$$J_2(\eta) = \iint_A 2\omega(x, y, \eta, \eta_x, \eta_y)\, dx ,$$

where

$$2\omega = f_{zz}\,\eta^2 + 2f_{zp}\,\eta\eta_x + 2f_{zq}\,\eta\eta_y + f_{pp}\,\eta_x^2 + 2f_{pq}\,\eta_x\eta_y + f_{qq}\,\eta_y^2 .$$

A variation $\eta(x, y)$ will be termed *admissible* if it is a regular surface vanishing on the boundary C of A. The Euler equations for this integral are the equations

$$L(\eta) = \omega_\eta - (\partial/\partial x)\omega_{\eta_x} - (\partial/\partial y)\omega_{\eta_y} = 0 ,$$

and will be called the *accessory equations.* A solution of class C'' of these equations will be called an *accessory extremal.*

The quadratic functional $J_2(\eta)$ has associated with it a bilinear functional

$$J_2(\xi, \eta) = \iint_A \Omega(\xi, \eta)\, dx ,$$

where

$$\Omega(\xi, \eta) = \xi\omega_\eta + \xi_x\omega_{\eta_x} + \xi_y\omega_{\eta_y} .$$

If the function $\xi(x, y)$ is of class C'' and vanishes on the boundary C of A, the condition

$$J_2(\xi, \eta) = 0$$

will be called a *natural isoperimetric condition* for the functional $J_2(\eta)$. It is satisfied by every accessory extremal $\eta(x, y)$, since by the use of Green's theorem it is found that in this case

$$J_2(\xi, \eta) = \int_C \xi(\omega_{\eta_x}\, dy - \omega_{\eta_y}\, dx) = 0 .$$

Conversely, a variation $\eta(x, y)$ of class C'' which satisfies all natural isoperimetric conditions is an extremal surface. This follows from the fundamental lemma, since now we have

$$J_2(\xi, \eta) = \iint_A L(\eta)\xi \, dx\, dy + \int_C \xi(\omega_{\eta_x} dy - \omega_{\eta_y} dx) = \iint_A L(\eta)\xi \, dx\, dy = 0$$

for every set of admissible variations ξ of class C''.

Consider now a set of natural isoperimetric conditions

$$J_2(\xi_\beta, \eta) = 0 \qquad (\beta = 1, \cdots, m). \tag{30.1}$$

The Euler equations of J_2 subject to the conditions (30.1) are the equations

$$L(\eta + \mu_\beta \xi_\beta) = 0, \qquad \mu_\beta' = 0.$$

A solution η, μ of these equations of class C'' will be called an *isoperimetric extremal.*

The set of conditions (30.1) will be called a proper set of natural isoperimetric conditions if the determinant

$$| J_2(\xi_\alpha, \xi_\beta) | \qquad (\alpha, \beta = 1, \cdots, m)$$

is different from zero.

Lemma 30.1. *If the set* (30.1) *is a proper set of natural isoperimetric conditions, every isoperimetric extremal η_i, μ_β satisfying these conditions has $\mu_\beta = 0$.*

For, by the use of Green's theorem, and from the fact that $\xi_\alpha \equiv 0$ on the boundary C of A, it is found that

$$J_2(\xi_\alpha, \eta) = \iint_A \xi_\alpha L(\eta)\, dx\, dy = -\mu_\beta \iint_A \xi_\alpha L(\xi_\beta)\, dx\, dy = -J_2(\xi_\alpha, \xi_\beta)\mu_\beta = 0,$$

and hence that $\mu_\beta = 0$, as was to be proved.

The set of conditions (30.1) will be called a *minimal set* for $J_2(\eta)$ if the inequality $J_2(\eta) \geqq 0$ is true for every admissible variation $\eta(x, y)$ satisfying the conditions (30.1), and if no proper sub-set of these conditions has this property.

Theorem 30.1. *If the set* (30.1) *is a minimal set, there exists a proper minimal set composed of the same number of natural isoperimetric conditions.*

This result can be proved by the methods in §3 used in the fixed end point case for simple integrals. In a similar manner we obtain further results.

Theorem 30.2. *If the set* (30.1) *forms a proper minimal set, the inequality*

$$J_2(\xi_\alpha, \xi_\beta)a_\alpha a_\beta < 0 \qquad (\alpha, \beta = 1, \cdots, m) \tag{30.2}$$

is true for every set of constants $(a) \neq (0)$. *Conversely, if the set* (30.1) *forms a maximal set of natural isoperimetric conditions such that the conditions* (30.2) *hold, the set* (30.1) *forms a minimal set.*

Theorem 30.3. *The number of conditions in a minimal set of natural isoperimetric conditions is the same for every such set.*

In view of the last theorem we may define the *type number* of the functional $J_2(\eta)$ to be equal to the number of isoperimetric conditions in a minimal set, if such a set exists. If there is no minimal set, the type number of $J_2(\eta)$ will be said to be infinite.

THEOREM 30.4. *If the set* $\eta_\alpha(x, y)$ $(\alpha = 1, \cdots, m)$ *forms a set of* m *admissible variations satisfying the conditions*

$$J_2(\eta_\alpha, \eta_\beta)a_\alpha a_\beta < 0 \qquad (\alpha, \beta = 1, \cdots, m) \tag{30.3}$$

for every set of constants $(a) \neq (0)$, *the type number of* J_2 *is at least* m. *If these functions form a maximal set of such variations, the type number of* J_2 *is equal to* m.

This follows at once from Theorem 30.2, since the functions $\eta_\alpha(x, y)$ can be replaced by admissible variations $\xi_\alpha(x, y)$ of class C'' without disturbing the inequality (30.3).

The *condition of Legendre* on a surface S will be said to hold if the inequality

$$Q(x, y, \theta) = f_{pp} \cos^2 \theta + 2f_{pq} \cos \theta \sin \theta + f_{qq} \sin^2 \theta \geqq 0$$

is true on S for every constant θ. If the inequality sign is excluded, the *strengthened condition of Legendre* will be said to be satisfied on S.

THEOREM 30.5. *If the type number of* $J_2(\eta)$ *is finite, the condition of Legendre holds on the extremal* S. *Moreover, if the expression* $f_{pp}f_{qq} - f_{pq}^2$ *is different from zero on* S, *the strengthened condition of Legendre holds on* S.

To prove this let m be the type number of $J_2(\eta)$, and suppose the theorem is false. Then at some point (x_0, y_0) and for a particular value θ_0 we could have $Q(x_0, y_0, \theta_0) < 0$. In fact, there would exist $m + 1$ points $P_\alpha \equiv (x_\alpha, y_\alpha)$ $(\alpha = 0, 1, \cdots, m)$ such that

$$Q(x_\alpha, y_\alpha, \theta_0) < 0 \qquad (\alpha = 0, 1, \cdots, m).$$

With each of these points P_α as a center we could describe a set of $m + 1$ circles h_α in A which do not intersect. By the methods used by Mason[32] we could construct a set of $m + 1$ admissible variations $\eta_\alpha(x, y)$ having $\eta_\alpha \equiv 0$ outside the circle h_α and having $J(\eta_\alpha) < 0$. It would follow that

$$J_2(\eta_\alpha, \eta_\alpha) < 0, \qquad J_2(\eta_\alpha, \eta_\beta) = 0 \qquad (\alpha \neq \beta; \alpha, \beta = 0, 1, \cdots, m),$$

and hence that the type number of $J_2(\eta)$ would be at least $m + 1$, by Theorem 30.4. From this contradiction we infer the truth of Theorem 30.5.

A continuous curve C composed of a finite number of regular sub-arcs will be called a *conjugate curve* if there exists an accessory extremal $\eta(x, y)$ having $\eta \equiv 0$ on C but not both $\eta_x \equiv 0$, $\eta_y \equiv 0$ on C.

We have the following interesting theorem concerning these curves.

THEOREM 30.6. *Suppose the strengthened condition of Legendre holds on the extremal surface* S. *If the type number* m *of the functional* $J_2(\eta)$ *on* S *is finite, the number of curves in any set of non-intersecting conjugate curves cannot exceed* m.

To prove this let C_α $(\alpha = 1, \cdots, p)$ be a set of non-intersecting conjugate curves and $\eta_\alpha(x, y)$ be the corresponding accessory extremals having $\eta_\alpha \equiv 0$ on C_α and $(\eta_{\alpha x}, \eta_{\alpha y}) \not\equiv (0, 0)$ on C_α. Let $P_\alpha \equiv (x_\alpha, y_\alpha)$ be a point on C_α at which $(\eta_{\alpha x}, \eta_{\alpha y}) \neq (0, 0)$. With P_α as a center, describe a circle h_α which does

[32] See Bolza, *Vorlesungen über Variationsrechnung*, pp. 673–5.

not intersect the curves $C_\beta(\beta \neq \alpha)$. Let $u_\alpha(x, y)$ be an admissible variation which is identically zero except on the circle h_α, and is such that

$$\int_{C_\alpha} u_\alpha[\omega_{\eta_x}(\eta_\alpha)\, dy - \omega_{\eta_y}(\eta_\alpha)\, dx] \neq 0\,. \tag{30.4}$$

Such a function surely exists, since otherwise the expression

$$\omega_{\eta_x}(\eta_\alpha)\, dy - \omega_{\eta_y}(\eta_\alpha)\, dx = (f_{pp}\,\eta_{\alpha x} + f_{pq}\,\eta_{\alpha y})\, dy - (f_{pq}\,\eta_{\alpha x} + f_{qq}\eta_{\alpha y})\, dx$$

would be identically zero, by virtue of the fundamental lemma in the calculus of variations. Moreover, on C_α we have

$$\eta_{\alpha x}\, dx + \eta_{\alpha y}\, dy = 0,$$

since $\eta_\alpha \equiv 0$ on C_α. But these relations and the Legendre condition would imply that $(\eta_{\alpha x}, \eta_{\alpha y}) = (0, 0)$ at P_α, which is not the case. The functions $u_\alpha(x, y)$ accordingly can be chosen as described.

As a next step in the proof of Theorem 30.6 we set $\bar{\eta}_\alpha = \eta_\alpha$ within C_α and identically zero elsewhere. By the use of Green's theorem it is seen that

$$\begin{aligned} J_2(\bar{\eta}_\alpha) &= \iint_{A_\alpha} \Omega(\eta_\alpha, \eta_\alpha)\, dx\, dy \\ &= \int_{C_\alpha} \eta_\alpha[\omega_{\eta_x}(\eta_\alpha)\, dy - \omega_{\eta_y}(\eta_\alpha)\, dx] = 0, \end{aligned}$$

where A_α denotes the region bounded by C_α. If now we set $\xi_\alpha = \bar{\eta}_\alpha + b_\alpha u_\alpha$ (α not summed), it is found that

$$\begin{aligned} J_2(\xi_\alpha) &= J_2(\bar{\eta}_\alpha) + 2b_\alpha J_2(\bar{\eta}_\alpha, u_\alpha) + b_\alpha^2 J_2(u_\alpha) \\ &= 2b_\alpha J_2(\bar{\eta}_\alpha, u_\alpha) + b_\alpha^2 J_2(u_\alpha) \qquad (\alpha \text{ not summed}). \end{aligned}$$

By the use of Green's lemma again one sees that $J_2(\bar{\eta}_\alpha, u_\alpha)$ is equal to the expression (30.4). We may accordingly choose b_α such that $J_2(\xi_\alpha) < 0$. We shall now prove that $J_2(\xi_\alpha, \xi_\beta) = 0$ $(\alpha \neq \beta)$. This result is clearly true if the curve C_β does not lie in the region A_α bounded by C_α. If C_β is within A_α we have

$$J_2(\xi_\alpha, \xi_\beta) = J_2(\xi_\alpha, \bar{\eta}_\beta) + h_\beta J_2(\xi_\alpha, u_\beta).$$

By the usual integration by parts it is seen that

$$\begin{aligned} J_2(\xi_\alpha, \bar{\eta}_\beta) &= \iint_{A_\beta} \Omega(\xi_\alpha, \eta_\beta)\, dx\, dy \\ &= \iint_{A_\beta} \eta_\beta L(\xi_\alpha)\, dx\, dy + \int_{C_\beta} \eta_\beta[\omega_{\eta_x}(\xi_\alpha)\, dy - \omega_{\eta_y}(\xi_\alpha)\, dx] = 0\,. \end{aligned}$$

Similarly, we have $J_2(\xi_\alpha, \eta_\beta) = 0$, and hence $J_2(\xi_\alpha, \xi_\beta) = 0$ in this case also. We have accordingly

$$J_2(\xi_\alpha, \xi_\alpha) < 0, \qquad J_2(\xi_\alpha, \xi_\beta) = 0 \qquad (\alpha \neq \beta; \alpha, \beta = 1, \cdots, p).$$

It follows from Theorem 30.4 that p cannot exceed the type number m of $J_2(\eta)$, as was to be proved.

It is clear that this last result can be extended to sets of conjugate curves which may have points or sub-arcs in common.

31. **The existence of minimal sets.** We shall now show that under certain conditions, which are known to be satisfied in many cases, the type number of the functional $J_2(\eta)$ is finite. We assume that the boundary C of A is regular, and that the strengthened condition of Legendre holds.

The accessory boundary value problem is given by the system

$$L(u) = \sigma u, \qquad u \equiv 0 \text{ on } C. \tag{31.1}$$

A value of σ for which this system has a solution $(u) \not\equiv (0)$ of class C'' on A will be called a *characteristic root*, and the corresponding function $u(x, y)$ will be called a *characteristic solution*.

Suppose now that there exists a sequence of characteristic roots

$$\sigma_1 \leqq \sigma_2 \leqq \cdots \leqq \sigma_n \leqq \cdots \tag{31.2}$$

with $\lim_{n\to\infty} \sigma_n = \infty$ and a sequence of corresponding characteristic solutions

$$u_1, u_2, \cdots, u_n, \cdots$$

such that every function $\eta(x, y)$ of class C'' with $\eta \equiv 0$ on C can be expressed by a uniformly convergent series in the form

$$\eta(x, y) = \sum_{k=1}^{\infty} c_k u_k, \tag{31.3}$$

where the c's are constants. We suppose further that

$$\iint_A u_i u_k \, dx \, dy = \delta_{ik} \qquad [\delta_{ii} = 1, \delta_{ik} = 0 \ (i \neq k)]. \tag{31.4}$$

An immediate consequence of this assumption is that

$$c_k = \iint_A \eta u_k \, dx \, dy, \tag{31.5}$$

as one readily verifies by multiplying the equation (31.3) by u_i and integrating term by term.

THEOREM 31.1. *Let m be the number of negative characteristic roots in the sequence* (31.2). *If we set $\xi_\alpha = u_\alpha$ $(\alpha = 1, \cdots, m)$, the inequality*

$$J_2(\xi_\alpha, \xi_\beta) a_\alpha a_\beta < 0 \qquad (\alpha, \beta = 1, \cdots, m) \tag{31.6}$$

holds for every set of constants $(a) \neq (0)$. *Moreover,* $J_2(\eta) \geqq 0$ *for every admissible variation* $\eta(x, y)$ *having* $\eta \equiv 0$ *on* C *and satisfying the conditions*

$$J_2(\xi_\alpha, \eta) = 0 \qquad (\alpha = 1, \cdots, m). \tag{31.7}$$

To prove this we first note that by integration by parts we have

$$J_2(\xi_\alpha, \eta) = \iint_A \eta L(\xi_\alpha)\, dx\, dy = \sigma_\alpha \iint_A \xi_\alpha \eta\, dx\, dy, \tag{31.8}$$

since $\xi_\alpha = u_\alpha$. It follows from equations (31.4) that

$$J(\xi_\alpha, \xi_\alpha) = \sigma_\alpha < 0, \qquad J_2(\xi_\alpha, \xi_\beta) = 0,$$

and hence that the condition (31.6) is satisfied.

Consider now a surface $\eta(x, y)$ of class C'' having $\eta \equiv 0$ on C and satisfying the conditions (31.7). Such a surface is expressible in the form (31.3). The first m constants in this expression are all zero, since

$$c_\alpha = \iint_A \eta \xi_\alpha\, dx\, dy = (1/\sigma_\alpha) J(\xi_\alpha, \eta) = 0 \qquad (\alpha \text{ not summed}).$$

If now we multiply the series (31.3) by $L(\eta)$ and integrate, we find that

$$\iint_A \eta L(\eta)\, dx\, dy = \sum_{k=m+1}^{\infty} c_k \iint_A u_k L(\eta)\, dx\, dy.$$

From this expression we see by integration by parts and the use of equations (31.1) and (31.5) that

$$J(\eta) = \sum_{k=m+1}^{\infty} \sigma_k c_k^2 \geqq 0.$$

The equality holds only in the case

$$\eta = c_{m+1} u_{m+1} + \cdots + c_{m+s} u_{m+s},$$

where $\sigma_{m+1} = \cdots = \sigma_{m+s} = 0$, $\sigma_{m+s+1} > 0$. This proves the theorem for surfaces η of class C''.

In order to show that the theorem holds for all admissible surfaces vanishing on C, we note that every such surface $\eta(x, y)$ is a limit of a sequence of admissible surfaces $\eta_k(x, y)$ of class C''. Let $\mu_{\beta k}$ be a set of constants satisfying the conditions

$$\sigma_\alpha \mu_{\alpha k} = J_2(\xi_\alpha, \xi_\beta)\mu_{\beta k} = J_2(\xi_\alpha, \eta_k).$$

From this equation we see that $\lim_{k \to \infty} \mu_{\beta k} = 0$ if η satisfies the conditions (31.7). It follows that the sequence $\eta_k - \xi_\beta \mu_{\beta k}$ satisfies the conditions (31.7), has

$$J_2(\eta_k - \xi_\beta \mu_{\beta k}) \geqq 0 \qquad (k = 1, 2, \cdots),$$

and has $\eta(x, y)$ as its limit. Hence $J_2(\eta) \geqq 0$, as was to be proved.

COROLLARY. *The set* (31.7) *forms a minimal set, and the type number of* $J_2(\eta)$ *is equal to* m.

32. **Natural isoperimetric conditions in the general case.** We shall now return to the functional J described in §29. If $\xi(x, y)$ is an admissible variation of class C'', the condition

$$J_1(\xi) = \iint_A \{\xi f_z + \xi_x f_p + \xi_y f_q\}\, dx\, dy = 0$$

will be called a *natural isoperimetric condition.* This equation is satisfied by every extremal S, as can be seen by the use of Green's theorem. Conversely, every admissible surface of class C'' which satisfies all natural isoperimetric conditions is necessarily an extremal surface. This result is readily obtained by the use of Green's theorem and the fundamental lemma for double integrals in the manner described in §30.

Consider now a set of m natural isoperimetric conditions

$$J_1(\xi_\beta) = 0 \qquad (\beta = 1, \cdots, m). \tag{32.1}$$

Such a set will be called a *minimal set* for an extremal S if S affords a weak minimum to the integral J relative to neighboring admissible surfaces satisfying the conditions (32.1), and if no proper sub-set of these conditions has this property. The conditions (32.1) will be said to form a *proper set of* natural isoperimetric conditions for S if the determinant

$$| J_2(\xi_\alpha, \xi_\beta) | \tag{32.2}$$

is different from zero on S. In this case there are infinitely many admissible surfaces satisfying the condition (32.1), as can be seen from the proof of the following theorem.

THEOREM 32.1. *If the set* (32.1) *forms a proper minimal set for an extremal surface* S, $J_2(\eta) \geqq 0$ *on* S *for every set of admissible variations* η *satisfying on* S *the conditions*

$$J_2(\xi_\alpha, \eta) = 0 \qquad (\alpha = 1, \cdots, m). \tag{32.3}$$

To prove this let $z(x, y)$ be the function defining the surface S, and $\eta(x, y)$ be an arbitrary admissible variation satisfying the conditions (32.3). When the function

$$z(x, y) + a_\alpha \xi_\alpha(x, y) + b\eta(x, y) \tag{32.4}$$

is substituted for z in $J_1(\xi_\alpha)$, a set of m functions $J_{1\alpha}(a, b)$ is obtained, for which $J_{1\alpha}(0, 0) = 0$. Moreover, the functional determinant of $J_{1\alpha}(a, b)$ with respect to the variables $a_1, \cdots, a_m$ at $(a, b) = (0, 0)$ is the determinant (32.2), and is accordingly different from zero. It follows that the equations

$J_{1\alpha}(a, b) = 0$ have solutions $a_\alpha = A_\alpha(b)$ of class C' with $A_\alpha(0) = 0$. By differentiating the identity

$$J_{1\alpha}[A(b), b] \equiv 0$$

with respect to b it is found that for $b = 0$

$$0 = J_2(\xi_\alpha, \xi_\beta)A'_\beta(0) + J_2(\xi_\alpha, \eta) = J_2(\xi_\alpha, \xi_\beta)A'_\beta(0),$$

and hence that $A'_\beta(0) = 0$. It is clear that the one-parameter family of admissible surfaces

$$z(x, y, b) = z(x, y) + A_\alpha(b)\xi_\alpha(x, y) + b\eta(x, y)$$

satisfies the equations (32.1), contains the extremal S for $b = 0$, and has $z_b(x, y, 0) = \eta(x, y)$. It follows that the function $J(b)$ obtained by substituting this family in the functional J satisfies the relation $J(b) \geqq J(0)$, since the set (32.1) forms a minimal set for S. We have, accordingly, $J'(0) = 0$, $J''(0) = J_2(\eta) \geqq 0$, as was to be proved.

THEOREM 32.2. *If the set* (32.1) *forms a proper minimal set for an extremal S, the condition of Legendre holds on S.*

This result follows at once from Theorems 32.1 and 30.5.

A surface S will be said to satisfy the *condition of Weierstrass* if at each element (x, y, z, z_x, z_y) on it the inequality

$$E(x, y, z, z_x, z_y, p, q) \geqq 0$$

is true for every admissible set (x, y, z, p, q), where

$$E = f(x, y, z, p, q) - f(x, y, z, z_x, z_y) - (p - z_x)f_p(x, y, z, z_x, z_y) - (q - z_x)f_q(x, y, z, z_x, z_y).$$

THEOREM 32.3. *Suppose the set* (32.1) *forms a proper minimal set for an extremal surface S. If S affords a strong minimum relative to neighboring admissible surfaces satisfying the conditions* (32.1), *the condition of Weierstrass holds on S.*

In the case that S forms a minimum in the classical sense the theorem is usually proved by showing that if the Weierstrass condition fails to hold at some point on S, there exists a one-parameter family of surfaces $\eta(x, y, b)$ which is well defined for $b \geqq 0$, contains S for $b = 0$, and is such that $J'(0) < 0$ for the function $J(b)$ obtained by substituting this family in the integral J.[33] We shall use this fact to establish the more general theorem here given.

By an argument like that used in the proof of Theorem 32.1 it is readily seen that there exists a set of functions $A_\alpha(b)$ of class C' for $b \geqq 0$ such that the family of admissible surfaces

$$z(x, y, b) = z(x, y) + A_\alpha(b)\xi_\alpha(x, y)$$

[33] Bliss, *loc. cit.*

is well defined for $b \geqq 0$, satisfies the conditions (32.1), and contains S for $b = 0$. If now we substitute the functions $z + a_\alpha \xi_\alpha$ for z in the integral J, a function $J(a, b)$ is obtained such that

$$J[A(b), b] \geqq J(0, 0)$$

for $b \geqq 0$ and sufficiently near zero. It follows that we must have

$$J_{a_\alpha}(0, 0)A'_\alpha(0) + J_b(0, 0) \geqq 0.$$

The values $J_{a_\alpha}(0, 0)$ are the values of the functions $J_1(\xi_\alpha)$ on S and hence are zero. Hence we have $J_b(0, 0) \geqq 0$. But $J_b(0, 0)$ is equal to the derivative $J'(0)$ of the function $J(b)$ described in the last paragraph and is accordingly negative. From this contradiction we infer the truth of Theorem 32.3.

HARVARD UNIVERSITY.

THE NEIGHBORHOOD OF A SEXTACTIC POINT ON A PLANE CURVE

By Ernest P. Lane

1. Introduction. In recent years projective differential geometers have manifested increasing interest in the neighborhoods of singular elements. For example, in studying the neighborhood of an ordinary point on an analytic plane curve, inflexion points are usually excluded from consideration as singular. However, Bompiani has constructed a theory[1] of the neighborhood of an inflexion point on a plane curve, which has found fruitful applications in some work[2] of Su, and also in a recent paper[3] of the author.

An inflexion point on a plane curve being defined as usual to be a point where the curve possesses a unique tangent having precisely three-point contact, the singularity which naturally presents itself next for consideration is *the sextactic point*, which is defined to be a point where the curve possesses a unique tangent having precisely two-point contact, and also possesses a proper osculating conic having precisely six-point contact with the curve. So at a sextactic point the osculating conic hyperosculates the curve in the same sense as does the inflexional tangent at an inflexion point.

In this note a brief study is made of the neighborhood of a sextactic point. In §2 a canonical power series expansion is deduced which represents an analytic plane curve in the neighborhood of a sextactic point on it. In §3 are found some applications of this expansion.

2. Canonical power series expansion. Let us establish a projective coördinate system in a plane, in which a point has non-homogeneous coördinates x, y and homogeneous coördinates x_1, x_2, x_3, connected by the relations $x = x_2/x_1$, $y = x_3/x_1$. The context will show in any instance which coördinates are being used. Then let us consider a curve C which, in the neighborhood of a point $O(b, c)$ on it, can be represented by a power series expansion of the form

$$y - c = a_1(x - b) + a_2(x - b)^2 + a_3(x - b)^3 + \cdots . \tag{1}$$

By suitable choice of the coördinate system this expansion can be very much simplified, and it is the purpose of this section to carry this simplification as far

Received June 6, 1935.

[1] E. Bompiani, *Per lo studio proiettivo-differenziale delle singularità*, Bollettino dell' Unione Matematica Italiana, vol. 5 (1926), p. 118.

[2] B. Su, *On certain quadratic cones projectively connected with a space curve and a surface*, Tôhoku Mathematical Journal, vol. 38 (1933), p. 233.

[3] E. P. Lane, *Plane sections through an asymptotic tangent of a surface*, Bulletin of the American Mathematical Society, vol. 41 (1935), p. 285.

as possible, the result being a canonical form for the expansion representing a plane curve in the neighborhood of a sextactic point.

It is well known[4] that if the vertex (1, 0, 0) of the triangle of reference is taken at the point O, then $b = c = 0$. Furthermore, if the side $y = 0$ of the triangle coincides with the tangent of the curve C at O, then $a_1 = 0$. If the vertex (0, 0, 1) is on the osculating conic of C at O, and if the side $x = 0$ is the polar line of the vertex (0, 1, 0) with respect to the osculating conic, then $a_3 = a_4 = 0$. Finally, if the unit point is on the osculating conic, then $a_2 = 1$. The equation (1), after these simplifications, becomes

$$y = x^2 + a_5x^5 + a_6x^6 + \cdots, \tag{2}$$

and the equation of the osculating conic is

$$y = x^2. \tag{3}$$

The coördinate system is now determined except for the actual position of the vertex (0, 1, 0) on the tangent, and that of the unit point on the osculating conic.

It is evident that *necessary and sufficient conditions that the point O be a sextactic point, as defined above, on the curve C are* $a_5 = 0$, $a_6 \neq 0$. Let us suppose from now on that these conditions are satisfied, and write $a_6 = a$, so that equation (2) becomes

$$y = x^2 + ax^6 + a_7x^7 + a_8x^8 + \cdots \qquad (a \neq 0). \tag{4}$$

It will be useful to consider the cubic curves having contact of various orders with the curve C at the point O. The equation of *the cubics having six-point contact* with C at O is found by writing the most general equation of the third degree in x, y and demanding that this equation be satisfied identically in x as far as the terms in x^5 by the power series (4) for y; the result can be written in the form

$$(y - x^2)(1 + Fx + Gy) + Hy^3 = 0, \tag{5}$$

where F, G, H are arbitrary constants. It is obvious geometrically that each one of this three-parameter family of six-point cubics has a sextactic point at O. Moreover, the tangent, $y = 0$, at O meets each of these cubics in the corresponding *tangential* of O, namely, the point $(-F, 1, 0)$. This point is an inflexion point on any one of the six-point cubics through it, the equation of the corresponding inflexional tangent being

$$1 + Fx + Gy = 0. \tag{6}$$

It is easy to show that the cubics (5) having seven-point contact with the curve C at the point O are characterized by the condition

$$H = -a, \tag{7}$$

[4] Lane, *Projective Differential Geometry of Curves and Surfaces*, University of Chicago Press, 1932, p. 14.

and that for the eight-point cubics the condition is

$$F = -a_7/a \,. \tag{8}$$

It follows that *the tangent,* $y = 0$, *meets all the eight-point cubics in the same tangential point* $(a_7, a, 0)$. But the point distinct from O at which all the eight-point cubics intersect is habitually called the *Halphen point* of O. Thus we have proved the following theorem.

The Halphen point corresponding to a sextactic point O of a plane curve C is the tangential of O with respect to every one of the one-parameter family of eight-point cubics of C at O.

The position of the vertex $(0, 1, 0)$ on the tangent, $y = 0$, will now be prescribed by taking it to be the Halphen point. Then $a_7 = 0$. Therefore $F = 0$, and *the equation of the eight-point cubics* becomes

$$(y - x^2)(1 + Gy) - ay^3 = 0 \,, \tag{9}$$

where G is an arbitrary constant. It is worthy of note that *the only eight-point cubic with a double point at the point O is composite, its two components being the tangent line and the osculating conic.* Moreover, *there is only one non-composite eight-point cubic with the property that its inflexional tangent,* $1 + Gy = 0$, *at the Halphen point is tangent to the osculating conic,* $y = x^2$; *the equation,*

$$y - x^2 - ay^3 = 0 \,, \tag{10}$$

of this cubic is obtained by setting $G = 0$ *in equation* (9). Finally, *the line* $x = 0$ *is the harmonic polar of the Halphen point* $(0, 1, 0)$ *with respect to every one of the eight-point cubics.*

Among the eight-point cubics (9) is *the osculating, or nine-point, cubic.* Demanding that equation (9) be satisfied identically in x as far as the terms in x^8 by the series (4) for y, with $a_7 = 0$, we find the value of the parameter G for the osculating cubic:

$$G = -a_8/a \,. \tag{11}$$

The equation, $a - a_8 y = 0$, of the inflexional tangent of the osculating cubic at the Halphen point shows that the following theorem is true.

The inflexional tangent of the osculating cubic at the Halphen point is tangent to the osculating conic if, and only if, $a_8 = 0$.

Let us suppose hereinafter that $a_8 \neq 0$. Then the inflexional tangent of the osculating cubic meets the osculating conic, $y = x^2$, in the two points

$$\left(\pm\left(\frac{a}{a_8}\right)^{\frac{1}{2}}, \frac{a}{a_8}\right).$$

The characterization of the coördinate system will now be completed by taking one of these points for the unit point. Then $a_8 = a$, and we have reached the desired canonical form of the equation (1). Some of the results of this section are contained in the following summary.

By suitable choice of the projective coördinate system the expansion representing a plane curve C in the neighborhood of a sextactic point O which is not a singularity of any higher type can be reduced to the form

$$y = x^2 + ax^6 + ax^8 + a_9x^9 + a_{10}x^{10} + \cdots \qquad (a \neq 0)\,. \tag{12}$$

The equations of certain elements are as follows.

The tangent of C at the sextactic point O$(0, 0)$ *is*

$$y = 0\,. \tag{13}$$

The osculating conic of C at O is

$$y = x^2\,. \tag{3}$$

The osculating cubic of C at O is

$$(y - x^2)(1 - y) - ay^3 = 0\,. \tag{14}$$

The inflexional tangent of the osculating cubic (14) *at the Halphen point* $(0, 1, 0)$ *is*

$$y = 1\,. \tag{15}$$

The lines from O to the intersections $(\pm 1, 1)$ *of the osculating conic* (3) *and the inflexional tangent* (15) *are*

$$y = \pm\, x\,. \tag{16}$$

3. The neighborhood of a sextactic point. The canonical expansion (12) and other results of the preceding section will now be used to study more intensively the neighborhood of a sextactic point on a plane curve.

First of all, by means of equation (14) it is easy to show that the equation of the polar conic of the sextactic point $O(0, 0)$ with respect to the osculating cubic (14) is

$$2y - x^2 - y^2 = 0\,. \tag{17}$$

It follows that *this conic passes through the points* $(\pm 1, 1)$ *previously characterized.*

By direct calculation the equation of the Hessian of the osculating cubic (14) is found to be

$$(1 - y)\,[y(1 + 3ay) - 2x^2 + (1 - y)^2] + x^2[1 + (3a - 1)y] = 0\,. \tag{18}$$

Hence the inflexional tangent, $y = 1$, of the osculating cubic at Halphen point $(0, 1, 0)$ is tangent to the Hessian at the point $(1, 0, 1)$ and intersects the Hessian also at the point $(0, 1, 0)$. The equation of the inflexional tangent of the Hessian at the Halphen point is

$$1 - (3a + 1)y = 0\,. \tag{19}$$

If we write the most general equation of the fourth degree in x, y lacking, however, the terms of the second and lower degrees, and if we then demand

that this equation be satisfied as far as the terms in x^{10} by the power series (12) for y, we find that *there exists a unique eleven-point quartic with a triple point at the sextactic point O.* In fact, *the equation of this quartic is found to be*

$$(20) \qquad (y^2 - x^2)(y - x^2) + ay^4 = 0 .$$

One branch of this curve passes through the point O tangent to the line $y = x$; another passes through O tangent to the line $y = -x$; the third branch has nine-point contact with the curve C at the point O, its tangent at O being, of course, the line $y = 0$.

Let us denote the left member of equation (20) by $Q(x, y)$, so that

$$(21) \qquad Q(x, y) = (y^2 - x^2)(y - x^2) + ay^4 .$$

If the power series (12) is substituted for y in the polynomial $Q(x, y)$, the result to terms of degree thirteen is

$$(22) \qquad Q(x, y) = -a_9x^{11} - (a_{10} - 4a^2 - a)x^{12} - (a_{11} - a_9)x^{13} + \cdots .$$

It is important to observe in subsequent calculations that *the eleven-point quartic* (20) *with a triple point at the sextactic point O of the plane curve C hyperosculates C if, and only if, $a_9 = 0$.* We shall suppose hereinafter that $a_9 \neq 0$.

Procedure analogous to that used in obtaining equation (20) enables one to calculate *the equation of the unique fourteen-point quintic with a quadruple point at the sextactic point O of the plane curve C*:

$$(23) \qquad a_9[(x^3 - xy^2 - a_9y^3/a)(y - x^2) - axy^4] + (a_{10} - 4a^2 - a)yQ(x, y) = 0 .$$

Let us now consider quartic curves with a double point at the sextactic point O. *The equation of the two-parameter family of eleven-point quartics with a double point at the sextactic point O of the plane curve C is*

$$(24) \qquad \begin{aligned} &K[x(1 - y - a_9x/a)(y - x^2) - axy^3 + a_9y^4] \\ &\qquad + L(y - x^2)^2 + MQ(x, y) = 0 , \end{aligned}$$

where K, L, M are arbitrary constants. If $K = L = 0$, $M \neq 0$, this equation reduces to equation (20). If $K = M = 0$, $L \neq 0$, we get the osculating conic counted twice. *The equation of the one-parameter family of twelve-point quartics with a double point at O is found to be*

$$(25) \qquad \begin{aligned} &K[(x - xy - a_9y^2/a)(y - x^2) - axy^3 + (a_{10} - 3a^2 - a)Q(x, y)/a_9] \\ &\qquad + L(y - x^2)^2 = 0 . \end{aligned}$$

If $L = 0$, $K \neq 0$, this is *the unique twelve-point quartic with the lines $x = 0$ and $y = 0$ for nodal tangents at the point O.* If $K = 0$, $L \neq 0$, we see that the *unique twelve-point quartic with coincident nodal tangents at the point O is composite, consisting of the osculating conic counted twice.* Finally, *the equation of the unique*

thirteen-point quartic with a double point at the sextactic point O of the plane curve C is $R(x, y) = 0$, where $R(x, y)$ is defined by placing

$$(26)\quad \begin{aligned} R(x,y) = {} & (x - xy - a_9y^2/a)\,(y - x^2) - axy^3 + (a_{10} - 3a^2 - a)Q(x, y)/a_9 \\ & - [a_{11} - 2a_9 - (a_{10} - 3a^2 - a)\,(a_{10} - 4a^2 - a)/a_9]\,(y - x^2)^2/a^2 . \end{aligned}$$

In fact, on substituting the power series (12) for y in the polynomial $R(x, y)$, we find, to terms of degree thirteen,

$$(27)\quad R(x, y) = Ax^{13} + \cdots ,$$

where A is defined by the formula

$$(28)\quad A = a_{12} - a_{10} - a_9^2/a - 4a^2 - (a_{11}/a_9 - 1)\,(a_{10} - 3a^2 - a) .$$

Let us now drop the restriction that each of the quartic curves under consideration must have a singular point at O. By direct calculation we find that *the equation of the one-parameter family of quartics having thirteen consecutive points in common with the curve C at the sextactic point O is*

$$(29)\quad \begin{aligned} H \Bigg\{ & \left[1 - \frac{a_9}{a}xy + \left(\frac{a_{11}}{a_9} - \frac{a_{10}}{a} + 2a - 1\right)y^2 - \left(\frac{a_{11}}{a_9} - 1\right)x^2\right](y - x^2) \\ & - ax^2y^2 + a\left(\frac{a_{11}}{a_9} - 2\right)y^4 - \left(\frac{A}{a^2} + \frac{a_{11}}{a_9} - 1\right)(y - x^2)^2\Bigg\} \\ & + KR(x, y) = 0 , \end{aligned}$$

where A is defined in (28) and $R(x, y)$ in (26), while H, K are arbitrary constants. Among the thirteen-point quartics (29) is *the osculating, or fourteen-point, quartic* for which the ratio of the constants H, K has the value given by

$$(30)\quad H[a_{13} + 2a_{11} - a_9(1 + a) - a_{11}^2/a_9 - 2\,a_9a_{10}/a] + AK = 0 .$$

Equation (27) shows that if $A = 0$ the quartic $R(x, y) = 0$ is the osculating quartic, which therefore has a node at O.

UNIVERSITY OF CHICAGO.

ON CRITERIA FOR FOURIER CONSTANTS OF L-INTEGRABLE FUNCTIONS OF SEVERAL VARIABLES

By Charles N. Moore

1. **Introduction.** In a note published in the Proceedings of the National Academy of Sciences I have given for Fourier constants of a function of one variable a criterion which includes several previously known criteria.[1] In a paper presented to the American Mathematical Society in December 1933 I obtained an analogous criterion for Fourier coefficients of functions of two variables.[2]

The criterion in the case of functions of one variable was also obtained independently by Cesari,[3] who more recently has formulated for the case of functions of one variable a more general criterion which includes the previous one and a variety of other interesting special cases.[4] It is the purpose of the present paper to obtain an analogous criterion for functions of several variables. For the sake of simplicity we shall give the detailed discussion only for the case of two variables, as the extension to the more general case is fairly obvious.

2. **Terminology.** We are given two doubly infinite sets of constants A_{mn}, C_{mn} $(m, n = 0, 1, 2, \cdots)$, formally related in the following manner

$$\sum_{m=0,\, n=0}^{\infty,\, \infty} A_{mn}x^my^n \sim 1 \Big/ \left[\sum_{m=0,\, n=0}^{\infty,\, \infty} C_{mn}x^my^n\right], \tag{1}$$

where the symbol $\sim$ indicates that like powers of x and y on the two sides of the relationship are to be set equal to each other. For ranges of values of x and y for which the two power series converge, the relationship becomes an equality.

Let us set

$$S_{mn}(x, y, p, q) = \sum_{i=0,\, j=0}^{i=p,\, j=q} C_{m-i,\, n-j}\beta_{ij}(x, y) \qquad (0 \leqq p \leqq m, 0 \leqq q \leqq n), \tag{2}$$

where

$$\begin{aligned} &\beta_{ij}(x, y) = \cos ix \cos jy \quad (i, j = 1, 2, \cdots), \qquad \beta_{00} = \tfrac{1}{4}, \\ &\beta_{0j}(x, y) = \tfrac{1}{2} \cos jy \quad (j = 1, 2, \cdots), \qquad \beta_{i0}(x, y) = \tfrac{1}{2} \cos ix \quad (i = 1, 2, \cdots). \end{aligned} \tag{3}$$

Received May 6, 1935; presented to the American Mathematical Society, October 27, 1934.

[1] Vol. 19(1933), pp. 846–848. Cf. also Bull. Amer. Math. Soc., vol. 39(1933), pp. 907–913.

[2] For abstract see Bull. Amer. Math. Soc., vol. 40(1934), p. 40.

[3] Annali d. R. Sc. di Pisa, (2), vol. 3(1934), pp. 105–134; see, in particular, pp. 119–129.

[4] L. Cesari, *Sulle condizioni sufficienti per le successioni di Fourier*, Boll. della Unione Mat. Ital., vol. 13(1934), pp. 100–104.

Suppose we have further two doubly infinite sets of constants, l_{mn}, λ_{mn}, which are positive for all m, n and satisfy

(4) $(1/l_{mn}) \mid S_{mn}(x, y, p, q) \mid < K \quad (\text{all } m, n, p, q; \; -\pi \leqq x \leqq y, -\pi \leqq y \leqq \pi),$

except for a set of points E of measure zero, and

$$(1/\lambda_{mn}) \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \mid S_{mn}(x, y, m, n) \mid dx\, dy < K_1 \qquad (\text{all } m, n), \tag{5}$$

where K and K_1 are positive constants.

Consider now any doubly infinite set of constants a_{mn}, such that a_{mn} tends to zero as m and n become infinite, or as either index becomes infinite while the other one remains fixed. We set

$$La_{mn} = \sum_{i=0,\ j=0}^{\infty,\ \infty} A_{ij} a_{m+i,\, n+j}\, , \tag{6}$$

it being now assumed that the A_{ij} are so chosen that the double series on the right hand side of (6) is convergent. We assume further that the La_{mn} are such that the series

$$\sum_{i=0,\ j=0}^{\infty,\ \infty} C_{ij} La_{m+i,\, n+j} \tag{7}$$

converges absolutely. Making use of (6), the series (7) may be rewritten in the form

$$\sum_{i=0,\ j=0}^{\infty,\ \infty} C_{ij} \sum_{p=0,\ q=0}^{\infty,\ \infty} A_{pq} a_{m+i+p,\, n+j+q}\, .$$

Let us now suppose that the terms of the above series may be rearranged in such a manner that all the terms involving a's with the same indices occur in groups, which is certainly the case, for example, if the quadruple series is absolutely convergent. If we set $i + p = r$, $j + q = s$, the rearranged series may be written

$$\sum_{r=0,\ s=0}^{\infty,\ \infty} \left(\sum_{i=0,\ j=0}^{i=r,\ j=s} C_{ij} A_{r-i,\, s-j} \right) a_{m+r,\, n+s}\, .$$

In view of (1) all the expressions in parentheses vanish, if either r or s exceeds zero, and for the case $r = 0 = s$, the expression has the value unity. Thus we obtain the relationship

$$a_{mn} = \sum_{i=0,\ j=0}^{\infty,\ \infty} C_{ij} La_{m+i,\, n+j}\, . \tag{8}$$

3. **The general theorem.** Our criterion for Fourier constants of functions of two variables may now be stated as follows:

THEOREM. *If we have a doubly infinite set of constants, a_{mn}, for which a_{mn} tends*

to zero as m and n become infinite, or as either index becomes infinite while the other remains fixed, and if these constants satisfy conditions

$$(9) \qquad \sum_{m=0,\, n=0}^{\infty,\, \infty} l_{mn}\, |La_{mn}| < \infty, \qquad \sum_{m=0,\, n=0}^{\infty,\, \infty} \lambda_{mn}\, |La_{mn}| < \infty,$$

where the l_{mn} and the λ_{mn} are the constants occurring in (4) *and* (5), *then the a_{mn} are the Fourier* cos-cos *coefficients of an L-integrable function.*

Proof. Making use of (8) and (2), we may make the following transformation of the $(p + 1)$ $(q + 1)$ terms in the upper left hand corner of the double trigonometric series whose general term is $a_{mn}\beta_{mn}(x, y)$:

$$\begin{aligned}
\sum_{i=0,\, j=0}^{p,\, q} a_{ij}\beta_{ij}(x, y) &= \sum_{i=0,\, j=0}^{p,\, q} \beta_{ij} \sum_{r=0,\, s=0}^{\infty,\, \infty} C_{rs} La_{i+r,\, j+s} \\
&= \sum_{m=0,\, n=0}^{p,\, q} \left(\sum_{i=0,\, j=0}^{m,\, n} C_{m-i,\, n-j}\beta_{ij} \right) La_{mn} \\
&\quad + \sum_{m=0,\, n=q+1}^{p,\, \infty} \left(\sum_{i=0,\, j=0}^{m,\, q} C_{m-i,\, n-j}\beta_{ij} \right) La_{mn} \\
(10) \qquad &\quad + \sum_{m=p+1,\, n=0}^{\infty,\, q} \left(\sum_{i=0,\, j=0}^{p,\, n} C_{m-i,\, n-j}\beta_{ij} \right) La_{mn} \\
&\quad + \sum_{m=p+1,\, n=q+1}^{\infty,\, \infty} \left(\sum_{i=0,\, j=0}^{p,\, q} C_{m-i,\, n-j}\beta_{ij} \right) La_{mn} \\
&= \sum_{m=0,\, n=0}^{p,\, q} S_{mn}(x, y, m, n)\, La_{mn} + \sum_{m=0,\, n=q+1}^{p,\, \infty} S_{mn}(x, y, m, q)\, La_{mn} \\
&\quad + \sum_{m=p+1,\, n=0}^{\infty,\, q} S_{mn}(x, y, p, n)\, La_{mn} + \sum_{m=p+1,\, n=q+1}^{\infty,\, \infty} S_{mn}(x, y, p, q)\, La_{mn}.
\end{aligned}$$

It follows from (4) and the first condition in (9) that the last three terms on the right hand side of (10) approach zero as p and q become infinite, throughout the region $(-\pi, \pi; -\pi, \pi)$ except for the set of points for which (4) does not hold. Also for the same reasons the first term approaches a definite limit as p and q become infinite, with the same exceptions. Passing to the limit, we obtain the result

$$(11) \qquad f(x, y) = \sum_{m=0,\, n=0}^{\infty,\, \infty} a_{mn}\beta_{mn}(x, y) = \sum_{m=0,\, n=0}^{\infty,\, \infty} S_{mn}(x, y, m, n)\, La_{mn} \qquad (E_1),$$

where the set E_1 over which $f(x, y)$ is defined is the set complementary to E with regard to the region $(-\pi, \pi; -\pi, \pi)$.

Let us now choose a set of ϵ_n's tending to zero as n becomes infinite, and represent by R_n a region which includes all points of the region $(-\pi, \pi; -\pi, \pi)$ except those in a set of regions of total area ϵ_n including the points of E. We then set $f_n(x, y) = |f(x, y)|$ in R_n and $= 0$ in the remaining part of $(-\pi, \pi; -\pi, \pi)$. Then we have from (11), the second part of (9), and (5)

$$
(12)\quad \begin{aligned}
\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} f_n(x, y)\, dx\, dy &= \iint_{R_n} |f(x, y)|\, dx\, dy \\
&\leq \sum_{m=0,\, n=0}^{\infty,\, \infty}\left[(1/\lambda_{mn})\iint_{R_n} |S_{mn}(x, y, m, n)|\, dx\, dy\right]\lambda_{mn}\,|La_{mn}| < K_2,
\end{aligned}
$$

where K_2 is a positive constant. Moreover, from the definition of $f_n(x, y)$, the above integral is a monotonic non-decreasing function of n. Hence its limit as n becomes infinite exists, and therefore by a well-known theorem[5] $|f(x, y)|$, the limit function of $f_n(x, y)$ for almost all points in $(-\pi, \pi; -\pi, \pi)$, is L-integrable in that region. Thus our theorem is proved.

4. **Special cases.** We consider now some of the more interesting special cases. Suppose the relationship (1) takes the form

$$(13)\qquad (1-x)^{r+1}(1-y)^{r+1} = [1/(1-x)^{-r-1}(1-y)^{-r-1}],$$

where $r > 0$. Then in (2) the coefficients on the right hand side are

$$(14)\qquad C_{mn}^{(r)} = \binom{r+m}{m}\binom{r+n}{n},$$

and

$$(15)\qquad S_{mn}(x, y, m, n)/C_{mn}^{(r)}$$

is the mean that arises in applying Cesàro summation of order r to the double trigonometric series whose terms are (3). This breaks up into the product of the means arising in applying the same process to the analogous simple series in x and y. From the well known expression for the partial sums of such series and the monotonic properties of the binomial coefficients used in forming the numerator of (15), it follows by an application of Abel's lemma that (15) is bounded for each (x, y) in $(-\pi, \pi; -\pi, \pi)$, except when one or the other of the variables is equal to zero. The same procedure serves to deal with the case where the arguments m and n in the numerator of (15) are replaced by $p < m$ and $q < n$. Moreover, in the various proofs concerning Cesàro summability of order r $(r > 0)$ of the Fourier series,[6] it is shown that the integral from $-\pi$ to π of the absolute value of the expression in one variable analogous to (15) is bounded for all n. Since the double integral of the absolute value of (15) over $(-\pi, \pi; -\pi, \pi)$ breaks up into the product of two simple integrals of the type indicated, it is readily seen that this double integral remains bounded for all m and n. It is apparent then that the l_{mn} and λ_{mn} of (4) and (5) may in the present case each be taken equal to (14). Since (14) is $O(m^r n^r)$ as m and n become infinite, and La_{mn} reduces to the double difference of order $(r + 1)$ of the a's when the

[5] Cf. Hobson, *Theory of Functions of a Real Variable*, 2nd or 3rd ed., vol. I, §399.

[6] For the simplest proof, see Marcel Riesz, *Sur la sommation des séries de Fourier*, Acta Lit. ac Sci., vol. 1 (1922–23), pp. 104–113; see, in particular, pp. 109–111.

A's are determined from the left hand side of (13), it follows that in the present case conditions (9) may be combined into

$$\Sigma\, m^r n^r \,|\, \Delta_{r+1,r+1} a_{mn} \,| < \infty. \tag{16}$$

We thus have the theorem referred to in the second footnote. A special case of this theorem is the extension to functions of two variables[7] of a criterion due to Young.

Let us now consider the case where the r in (14) is equal to zero. Then the C_{mn} are each equal to unity and (2) takes the form of a partial sum for the double trigonometric series whose terms are (3). If in (5) we take $\lambda_{mn} = \log m \log n$ ($m \geqq 3$, $n \geqq 3$), replacing $\log m$ or $\log n$ by unity when m or n is less than three, and take $l_{mn} = 1$ for all values of m and n, (4) and (5) are readily seen to be fulfilled from known facts about the partial sums of the ordinary Fourier series.[8] Conditions (9) may then be expressed in the form

$$\Sigma \log (m+2) \log (n+2) \,|\, \Delta a_{mn} \,| < \infty. \tag{17}$$

We thus have a generalization to functions of two variables of a criterion for functions of one variable due to Szidon. This generalization was contained in the paper referred to in footnote 7.

5. **Extension to n variables.** For the case of n-tuply infinite sets of constants and functions of n variables the expression $La_{m_1 m_2 \cdots m_n}$ is defined in the following manner

$$La_{m_1 m_2 \cdots m_n} = \sum_{i_1=0,\cdots,i_n=0}^{\infty,\cdots,\infty} A_{i_1 \cdots i_n} a_{i_1+m_1,\cdots,i_n+m_n}, \tag{18}$$

where the A's are the constants occurring in the analogue of (1), the C's in this analogue being used in the definition of $S_{m_1 \cdots m_n}(x_1 \cdots x_n;\, p_1 \cdots p_n)$. Conditions (9) take the form

$$\Sigma\, l_{m_1 \cdots m_n} \,|\, La_{m_1 \cdots m_n} \,| < \infty, \qquad \Sigma\, \lambda_{m_1 \cdots m_n} \,|\, La_{m_1 \cdots m_n} \,| < \infty, \tag{19}$$

where the l's and the λ's are the constants occurring in the analogues of (4) and (5).

THE INSTITUTE FOR ADVANCED STUDY AND UNIVERSITY OF CINCINNATI.

[7] *Proc. of the International Math. Cong. at Zürich*, 1932, vol. 2, pp. 121–122.
[8] Cf. Tonelli, *Serie Trigonometriche*, p. 263, for the proof that (5) is satisfied.

ON THE REPRESENTATION OF A POLYNOMIAL IN A GALOIS FIELD AS THE SUM OF AN ODD NUMBER OF SQUARES

By Leonard Carlitz

1. Introduction. Let $GF(p^n)$ denote a Galois field of order p^n, where p is an *odd* prime, and n is an arbitrary positive integer; let $\mathfrak{O}(x, p^n)$ denote the totality of polynomials in an indeterminate x with coefficients in $GF(p^n)$. We consider the problem of determining the number of representations of a polynomial in $\mathfrak{O}$ as a sum of squares of polynomials in $\mathfrak{O}$ satisfying certain restrictions. The case of an even number of squares has been treated elsewhere;[1] in the present paper, we consider the case of an odd number of squares. Certain results derived in the paper on the even case will be required in the discussion of the odd case.

Our problem may be described more precisely thus. Let $\alpha_1, \alpha_2, \cdots, \alpha_{2s+1}$ be $2s + 1$ non-zero elements of $GF(p^n)$; let L be a *primary* polynomial, that is, one in which the coefficient of the highest power of x is 1. Then

(A) if L is of even degree $2k$, and

$$\epsilon = \alpha_1 + \alpha_2 + \cdots + \alpha_{2s+1} \neq 0,$$

we seek the number of solutions of

$$\epsilon L = \alpha_1 X_1^2 + \alpha_2 X_2^2 + \cdots + \alpha_{2s+1} X_{2s+1}^2$$

in primary X_i each of degree k.

(B) If L is of arbitrary degree l, $2k$ any even integer $> l$, α any non-zero element of the Galois field, and

$$\alpha_1 + \alpha_2 + \cdots + \alpha_{2s+1} = 0,$$

we seek the number of solutions of

$$\alpha L = \alpha_1 X_1^2 + \alpha_2 X_2^2 + \cdots + \alpha_{2s+1} X_{2s+1}^2$$

in primary X_i each of degree k.

The number of solutions of problems (A) and (B) is expressed in terms of the Artin numbers[2] σ_j now to be defined. If Δ and M are in $\mathfrak{O}$, M primary, the symbol (Δ/M) will denote the generalized quadratic character of Δ with respect to M:

Received June 4, 1935.

[1] L. Carlitz, Transactions of the American Mathematical Society, vol. 35 (1933), pp. 397–410; cited as Representations.

[2] E. Artin, Mathematische Zeitschrift, vol. 19 (1924), pp. 153–246.

$$\left(\frac{\Delta}{1}\right) = 1\,, \qquad \left(\frac{\Delta}{M}\right) = 0 \qquad \text{for} \qquad (\Delta, M) \neq 1\,,$$

while for $(\Delta, M) = 1$,

$$\left(\frac{\Delta}{M}\right) = \left(\frac{\Delta}{P_1}\right) \cdots \left(\frac{\Delta}{P_t}\right), \qquad \text{where} \qquad M = P_1 \cdots P_t\,;$$

here (Δ, M) denotes the "greatest" common divisor of Δ and M. Then

$$\sigma_j = \sigma_j(\Delta) = \sum_{\deg G = j} \left(\frac{\Delta}{G}\right),$$

summed over all primary G of degree j. It follows by the quadratic reciprocity theorem that $\sigma_j = 0$ for $j \geqq$ the degree of Δ, ($\Delta \neq \beta M^2$, β in Galois field).

The final formulas are considerably simpler when L is "quadratfrei", that is, not divisible by the square of any polynomial of degree $\geqq 1$. If we suppose L quadratfrei, and for problem (A) put

$$\vartheta = (-1)^s \epsilon\, \alpha_1 \alpha_2 \cdots \alpha_{2s+1}\,,$$

while for problem (B) we put

$$\vartheta = (-1)^s \alpha\, \alpha_1 \alpha_2 \cdots \alpha_{2s+1}\,,$$

then for $\Delta = \vartheta L$, σ_j (Δ) as defined above, we find that the number of solutions is given by

$$p^{nk(s-1)} \{\sigma_k + p^{ns}\sigma_{k-1} + (p^{2ns} - p^n)\sigma_{k-2} + \cdots + (p^{nks} - p^{n(ks-2s+1)})\sigma_0\}\,.$$

This formula holds in either case, but for problem (B) may admit considerable simplification. Thus, for example, if N^k denotes the number of solutions of problem (B), and if in addition k is greater than the degree of L, the following recursion formula holds:

$$N^{k+1} = p^{n(2s-1)}\, N^k\,.$$

When L is not quadratfrei, the general formula for the number of solutions is rather complicated (see Theorem 8.1). For particular types of non-quadratfrei L, however, the final formulas are less elaborate. Thus the case L = a square leads to a fairly simple formula (see Theorem 8.2). Finally, in the case of problem (B), the assumption $k > 1$ leads to the theorem that, if $L = L_0 M^2$, L_0 quadratfrei, M of degree m, then

$$N^k(L) = N^{k-2m}(L_0)\, \delta(M)\,,$$

where $\delta(M)$ denotes a divisor function of M. In particular, if M also is quadratfrei, and $(L_0, M) = 1$, then $\delta(M)$ reduces to

$$\prod_{P \mid M} \left\{1 - \left(\frac{\Delta}{P}\right) | P |^{s-1} + | P |^{2s-1}\right\}.$$

We remark that the results of the present paper are slightly more general than the corresponding results on an even number of squares. In the present case, no restrictions on the coefficients α_i need be made. (Cf. §3).

When $L = 1$, some of the results of this paper require modification. We shall therefore assume, unless the contrary is stated, that L is always of degree $\geqq 1$.

2. **Number of solutions of quadratic congruences.** We shall require certain theorems on the number of solutions of quadratic congruences in a single unknown. Consider the congruence

$$X^2 \equiv \Delta \pmod{M}, \tag{2.01}$$

where Δ and M are arbitrary polynomials. It is convenient to assume M primary. We shall use the symbol $\nu(\Delta, M)$ to denote the number of incongruent solutions of (2.01); clearly, $\nu(\Delta, 1) = 1$. Also as above let the symbol (M, N) denote the greatest common divisor of M and N; to make it unique, we suppose it primary. Similarly, we define $(M, N)_2$ as the greatest (primary) square dividing both M and N. In other words,

$$(M, N) = Q \cdot (M, N)_2,$$

where Q is not divisible by the square of any polynomial (of degree $\geqq 1$).

To begin with, we have the following lemmas.

Lemma 1. *If $(M, N) = 1$, then*

$$\nu(\Delta, MN) = \nu(\Delta, M)\, \nu(\Delta, N). \tag{2.02}$$

Lemma 2. *If $(\Delta, M)_2 = E^2$, where E is of degree e, then*

$$\nu(\Delta, M) = P^{ne} \cdot \nu\left(\frac{\Delta}{E^2}, \frac{M}{E^2}\right). \tag{2.03}$$

There is no difficulty about Lemma 1. As for Lemma 2, from $(\Delta, M)_2 = E^2$ it is evident that (2.01) implies $E \mid X$. Put $X = EX_1$, $\Delta = E^2\Delta_1$, $M = E^2M_1$; then (2.01) becomes

$$X_1^2 \equiv \Delta_1 \pmod{M_1}. \tag{2.04}$$

By definition, this congruence has $\nu(\Delta_1, M_1)$ incongruent solutions (mod M_1). But each solution of (2.04) gives rise to p^{ne} solutions of (2.01), namely,

$$X \equiv (X_1 + M_1\Gamma)E \pmod{M},$$

where Γ runs through a complete residue system (mod E). The lemma now follows immediately.

When (2.02) is used, it is evidently necessary to calculate $\nu(\Delta, M)$ only in the case $M =$ a power of an irreducible polynomial: $M = P^e$, say. For $e = 1$, we have at once

$$\nu(\Delta, P) = 1 + \left(\frac{\Delta}{P}\right), \tag{2.05}$$

where it is understood that $(\Delta/P) = 0$ for $P \mid \Delta$. In the next place, for $P \nmid \Delta$, to each solution of

$$X^2 \equiv \Delta \pmod{P^e}$$

corresponds a unique solution of

$$Y^2 \equiv \Delta \pmod{P^{e+1}},$$

and conversely. This may be seen by writing $Y \equiv X + P^eZ \pmod{P^{e+1}}$. We have therefore

LEMMA 3. *For* $e \geqq 1$, $P \nmid \Delta$,

$$\nu(\Delta, P^e) = 1 + \left(\frac{\Delta}{P}\right). \tag{2.06}$$

In order to evaluate $\nu(\Delta, P^e)$ when $P \mid \Delta$, we make use of Lemma 2. There are several cases to consider. Let $P^\lambda \mid \Delta$, $P^{\lambda+1} \nmid \Delta$. If, first, $e \leqq \lambda$, then by Lemma 2

$$\nu(\Delta, P^e) = p^{nf\pi}\cdot\nu(\Delta P^{-2f}, P^{e-2f}),$$

where $e = 2f$ or $2f + 1$, and π is the degree of P. In case $e = 2f$, we have at once

$$\nu(\Delta, P^e) = p^{nf\pi}. \tag{2.07}$$

If $e > 2f$, then also $\lambda > 2f$, so that $P \mid \Delta P^{-2f}$, and therefore

$$\nu(\Delta P^{-2f}, P) = \nu(0, P) = 1,$$

by (2.05). In both cases, therefore, (2.07) holds.

Consider now the case $e > \lambda$. If λ is odd, $\lambda = 2\mu + 1$, say, then by Lemma 2

$$\nu(\Delta, P^e) = p^{n\mu\pi}\cdot\nu(\Delta\, P^{-2\mu}, P^{e-2\mu}).$$

Now $e - \mu \geqq 2$; since $\Delta P^{-2\mu}$ is divisible by only the first power of P, the congruence

$$X^2 \equiv \Delta P^{-2\mu} \pmod{P^{e-2\mu}}$$

is plainly impossible, so that

$$\nu(\Delta, P^e) = 0 \qquad (\lambda = 2\mu + 1 < e). \tag{2.08}$$

If λ is even, $\lambda = 2\mu$, say, then by Lemma 2,

$$\nu(\Delta, P^e) = p^{n\mu\pi}\cdot\nu(\Delta\, P^{-2\mu}, P^{e-2\mu}).$$

Since $P^{2\mu+1} \nmid \Delta$, we may use (2.06), from which it follows that

$$\nu(\Delta, P^e) = P^{n\mu\pi}\cdot\left\{1 + \left(\frac{\Delta P^{-2\mu}}{P}\right)\right\} \qquad (\lambda = 2\mu < e). \tag{2.09}$$

Combining (2.07), (2.08), (2.09), we have

LEMMA 4. *If* $P^\lambda \mid \Delta$, $P^{\lambda+1} \nmid \Delta$, *$P$ of degree* π, *then*

(i) $$\nu(\Delta, P^e) = p^{nf\pi}$$

for $e \leqq \lambda$, $e = 2f$ *or* $2f + 1$;

(ii) *for* $e > \lambda$,

$$\nu(\Delta, P^e) = \begin{cases} 0 & \text{for } \lambda \text{ odd}, \\ P^{n\mu\pi}\left[1 + \left(\dfrac{\Delta P^{-\lambda}}{P}\right)\right] & \text{for } \lambda = 2\mu . \end{cases}$$

We shall also need the following

LEMMA 5. *The number of primary polynomials J of degree j satisfying the congruence*

$$J^2 \equiv \Delta \pmod{M},$$

where $j \geqq m = \deg M$, *is*

$$\nu_j(\Delta, M) = p^{n(j-m)} \cdot \nu(\Delta, M) . \tag{2.10}$$

The lemma follows at once by putting $J = X + MH$, where H is primary of degree $j - m$, and X is a solution of (2.01) of degree $< m$. Note in particular that

$$\nu_m(\Delta, M) = \nu(\Delta, M) .$$

3. **A lemma.** As stated in the Introduction, the results on representation by an odd number of squares are slightly more general than those by an even number. This follows from a lemma whose analog for the even case is false:

LEMMA 3.1. *If* $\alpha_1, \alpha_2, \cdots, \alpha_{2s+1}$ *are arbitrary non-zero elements of* $GF(p^n)$, *they may be so numbered that*

$$(\alpha_1 + \alpha_2)(\alpha_3 + \alpha_4) \cdots (\alpha_{2s-1} + \alpha_{2s}) \neq 0 . \tag{3.01}$$

For $s = 1$, the assumption $\alpha_2 + \alpha_3 = \alpha_3 + \alpha_1 + \alpha_1 + \alpha_2 = 0$ implies $\alpha_1 = \alpha_2 = \alpha_3$, whence $2\alpha_1 = 0$. Since $p \neq 2$, $\alpha_1 = 0$, contrary to hypothesis.

Assume now that the lemma holds for the value s; in other words the truth of (3.01). Then the assumption

$$\beta(\alpha_{2s+1} + \alpha_{2s+2}) = \beta(\alpha_{2s+2} + \alpha_{2s+3}) = \beta(\alpha_{2s+3} + \alpha_{2s+1}) = 0 ,$$

where β stands for the left member of (3.01), implies $\alpha_{2s+1} = \alpha_{2s+2} = \alpha_{2s+3}$, whence $2\alpha_{2s+1} = 0$, $\alpha_{2s+1} = 0$, contrary to hypothesis. This completes the induction.

That the lemma is false in the case of an even number of terms is clear from the example $(1, 1, 1, -1)$. However, $(1, 1, 1, -1, -1)$ comes under the lemma; note that in this instance $\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = 0$. The point of this remark will appear in the next section (see (4.16) ff.).

4. **Preliminary formulas for problem** (A). Consider now the number of solutions of

$$\epsilon L = \alpha_1 X_1^2 + \dots + \alpha_{2s+1} X_{2s+1}^2 , \tag{4.01}$$

in primary X_i of degree k, where

$$\alpha_1 \alpha_2 \cdots \alpha_{2s+1} \neq 0 , \qquad \epsilon = \alpha_1 + \alpha_2 + \cdots + \alpha_{2s+1} \neq 0 ,$$

and L is primary of degree $2k$. By the lemma of §3 we may assume the α_i so numbered that (3.01) holds. However, as the example at the end of §3 indicates, $\gamma = \alpha_1 + \cdots + \alpha_{2s}$ may vanish. We assume first that $\gamma \neq 0$. If we write δ in place of α_{2s+1}, then evidently the number of solutions of (4.01) is given by

$$\sum_{\epsilon L = \gamma F + \delta X^2} n(\gamma F = \alpha_1 X_1^2 + \cdots + \alpha_{2s} X_{2s}^2) , \tag{4.02}$$

where the summation is taken over all primary F, X of degrees $2k$ and k, respectively, such that $\epsilon L = \gamma F + \delta X^2$, and the summand denotes the number of sets of primary X_i of degree k satisfying

$$\gamma F = \alpha_1 X_1^2 + \cdots + \alpha_{2s} X_{2s}^2 . \tag{4.03}$$

By virtue of (3.01), the number of solutions of (4.03) may be written down at once. We define the functions

$$\begin{aligned} \rho_t(F) &= \left(1 - \frac{1}{p^{nt}}\right) \sum_{M \mid F}^{m > k} | M |^t + \sum_{M \mid F}^{m = k} | M |^t , \\ \omega_t(F) &= \left(1 + \frac{1}{p^{nt}}\right) \sum_{M \mid F}^{m > k} (-1)^m | M |^t + \sum_{M \mid F}^{m = k} (-1)^m | M |^t , \\ &\qquad (m = \deg M , \ | M | = p^{nm}) , \end{aligned} \tag{4.04}$$

where in each instance the first summation is taken over all primary M dividing F, and of degree $> k$; the second summation is over primary M dividing F, and of degree $= k$. Then[3]

$$n(\gamma F = \alpha_1 X_1^2 + \cdots + \alpha_{2s} X_{2s}^2) = \rho_{s-1}(F) \text{ or } \omega_{s-1}(F) , \tag{4.05}$$

according as

$$\zeta = (-1)^s \alpha_1 \alpha_2 \cdots \alpha_{2s} \tag{4.06}$$

is a square or a non-square in $GF(p^n)$.

Let $\chi = \chi(\zeta) = 1$ or -1 according as ζ is or is not a square in $GF(p^n)$. Then (4.05) becomes

$$\begin{aligned} & n(\gamma F = \alpha_1 X_1^2 + \cdots + \alpha_{2s} X_{2s}^2) \\ &= \left(1 - \frac{\chi}{p^{n(s-1)}}\right) \sum_{M \mid F}^{m > k} \chi^m | M |^{s-1} + \sum_{M \mid F}^{m = k} \chi^m | M |^{s-1} . \end{aligned} \tag{4.07}$$

[3] Representations, Theorem 4.

Substituting into (4.02) we see that the number of solutions of (4.01) is

$$(4.08)\qquad \left(1 - \frac{\chi}{p^{n(s-1)}}\right) \sum_{\epsilon L = \gamma MU + \delta X^2}^{m>k} \chi^m \mid M \mid^{s-1} + \sum_{\epsilon L = \gamma MU + \delta X^2}^{m=k} \chi^m \mid M \mid^{s-1},$$

where the first summation is taken over all primary M, U, X of respective degrees m, u, k, such that $m + u = 2k$, $m > k$ and $\epsilon L = \gamma MU + \delta X^2$; in the second summation, $m = u = k$.

The first summation may be written in the form

$$(4.09)\qquad \sum_{U}^{u<k} \chi^u p^{n(2k-u)(s-1)} \sum_{\epsilon L = \gamma MU + \delta X^2} 1,$$

where the outer sum is over all primary U of degree $< k$. In the inner sum, U remains fixed; hence the latter sum is the number of primary X of degree k such that

$$X^2 \equiv \frac{\epsilon}{\delta} L \pmod{U}.$$

Lemma 2.5 evidently applies here: the number in question is therefore

$$\nu_k\left(\frac{\epsilon}{\delta} L, U\right) = p^{n(k-u)} \cdot \nu\left(\frac{\epsilon}{\delta} L, U\right).$$

Therefore (4.09) becomes

$$(4.10)\qquad \sum_{U}^{u<k} \chi^u p^{n(2k-u)(s-1)} p^{n(k-u)} \cdot \nu\left(\frac{\epsilon}{\delta} L, U\right).$$

In exactly the same way the second summation in (4.08) may be thrown into the form (4.10); however, the sum is now restricted to U of degree $= k$. Hence (4.08) becomes

$$(4.11)\qquad \begin{aligned} &p^{nk(s-1)} W_k\left(\frac{\epsilon}{\delta} L\right) \\ &\quad + \left(1 - \frac{\chi}{p^{n(s-1)}}\right) \sum_{u=0}^{k-1} \chi^u p^{n(2k-u)(s-1)} p^{n(k-u)} W_u\left(\frac{\epsilon}{\delta} L\right), \end{aligned}$$

where for brevity we put

$$(4.12)\qquad W_u(\Delta) = \sum_{\deg U = u} \nu(\Delta, U),$$

summed over all primary U of degree u. This completes the proof of

THEOREM 4.1. *If $\alpha_1, \cdots, \alpha_{2s+1}$ are non-zero elements of $GF(p^n)$, so numbered that*

$$(4.13)\qquad (\alpha_1 + \alpha_2) \cdots (\alpha_{2s-1} + \alpha_{2s}) \neq 0;$$

and if in addition

$$\gamma = \alpha_1 + \cdots + \alpha_{2s} \not\equiv 0\,, \qquad \epsilon = \gamma + \alpha_{2s+1} \not\equiv 0\,, \qquad \delta = \alpha_{2s+1}\,,$$

the number of solutions of (4.01) *is furnished by formula* (4.11).

We now consider the case $\gamma = 0$. It will be necessary to make use of the following theorem:[4]

The number of sets of primary X_i of degree k satisfying

$$\alpha F = \alpha_1 X_1^2 + \cdots + \alpha_{2s} X_{2s}^2\,, \tag{4.14}$$

where F is primary of degree $f < 2k$, $\alpha \not\equiv 0$, $\alpha_1 + \cdots + \alpha_{2s} = 0$, and (4.13) *holds, is given by*

$$p^{n(2k-f)(s-1)}\,\rho_{s-1}^{k}(F) \qquad \text{or} \qquad p^{n(2k-f)}\,\omega_{s-1}^{k}(F)\,,$$

according as $(-1)^s\alpha_1\cdots\alpha_{2s}$ is or is not a square in $GF(p^n)$.

The functions ρ and ω are defined by

$$\rho_t^k(F) = \left(1 - \frac{1}{p^{nt}}\right)\sum_{M|F}^{m>f-k} |\,M\,|^t + \sum_{M|F}^{m=f-k} |\,M\,|^t,$$

$$(-1)^f\,w_t^k(F) = \left(1 + \frac{1}{p^{nt}}\right)\sum_{M|F}^{m>f-k} (-1)^m\,|\,M\,|^t + \sum_{M|F}^{m=f-k} (-1)^m\,|\,M\,|^t.$$

(If $f < k$, the second summation in each case is defined $= 0$.) Hence, making use of the symbol χ defined above, we may express the number of solutions of (4.14) by means of the single formula:

$$p^{n(2k-f)(s-1)}\left\{\left(1 - \frac{\chi}{p^{n(s-1)}}\right)\sum_{M|F}^{m>f-k} \chi^{f-m}\,|\,M\,|^{s-1} + \sum_{M|F}^{m=f-k} \chi^{f-m}\,|\,M\,|^{s-1}\right\}. \tag{4.15}$$

Returning to (4.01), we see that, when $\gamma = 0$, the number of solutions in question is

$$\sum_{L=\Phi+X^2} n(\Phi = \alpha_1 X_1^2 + \cdots + \alpha_{2s} X_{2s}^2)\,, \tag{4.16}$$

where for convenience we put $\epsilon = \alpha_{2s+1} = 1$. The summation in (4.16) extends over *all* Φ of degree $< 2k$ such that $L - \Phi = X^2$. If L is a square, $\Phi = 0$ must be included. Let us assume first that L is not a square, so that we may put $\Phi = \alpha F$, say, where $\alpha \not\equiv 0$, and F is primary of degree $f < 2k$. Then the theorem concerning (4.13) may be applied. Hence substituting from (4.15) in (4.16), we see that the number of solutions of (4.01) is

$$\begin{aligned}
&\sum_{F}^{f<2k} p^{n(2k-f)(s-1)}\left\{\left(1 - \frac{\chi}{p^{n(s-1)}}\right)\sum_{\substack{L=\alpha F+X^2\\ F=MU}}^{u<k} \chi^u\,|\,M\,|^{s-1} + \sum_{\substack{L=\alpha F+X^2\\ F=MU}}^{u=k} \chi^u\,|\,M\,|^{s-1}\right\}\\
&\qquad = \left(1 - \frac{\chi}{p^{n(s-1)}}\right)\sum_{L=\alpha MU+X^2}^{u<k} \chi^u p^{n(2k-u)(s-1)} + \sum_{L=\alpha MU+X^2}^{u=k} \chi^u p^{nu(s-1)}.
\end{aligned} \tag{4.17}$$

[4] Representations, Theorem 5.

Now it is clear that for fixed U of degree $< k$,

$$\sum_{L=\alpha MU+X^2} 1 = \nu_k(L, U). \tag{4.18}$$

Hence exactly as in the derivation of (4.11), (4.17) becomes

$$\chi^k p^{nk(s-1)} W_k(L) + \left(1 - \frac{\chi}{p^{n(s-1)}}\right) \sum_{u=0}^{k-1} \chi^u p^{n(s-1)(2k-u)} p^{n(k-u)} W_u(L). \tag{4.19}$$

If now we recall that $\epsilon = \gamma + \alpha_{2s+1} = \alpha_{2s+1} = \delta$, we see that (4.19) is identical with (4.11). In other words, when L is not a square, (4.11) furnishes the number of solutions of (4.01) in the case $\gamma = 0$ also.

We consider finally the case L = a square. Since Φ in (4.16) may be 0, the above reasoning must be modified somewhat. Thus to (4.17) we must add the terms arising from $\Phi = 0$, in other words,

$$n(0 = \alpha_1 X_1^2 + \cdots + \alpha_{2s} X_{2s}^2; \deg X_i = k) \tag{4.20}$$

(the number of representations of 0 in the form $\alpha_1 X_1^2 + \cdots + \alpha_{2s} X_{2s}^2$, for which (4.13) holds). On the other hand, (4.18) is no longer valid, for the right member includes the possibility $X^2 = L$, excluded from the left member. But this is the condition that $\Phi = 0$. Hence in comparing (4.17) and (4.19), we see that the latter has been increased by 1 each time $\Phi = 0$. But the number of times this occurs is precisely (4.20). It follows, therefore, that in the present case also (4.19) furnishes the number of solutions of (4.01). Incidentally, we have evaluated (4.20). As we have just seen, for each U the right member of (4.18) contributes a unit to (4.20). Since there are p^{nu} distinct polynomials U of degree u, substitution in (4.17) gives rise to the following explicit formula for (4.20):

$$\left(1 - \frac{\chi}{p^{n(s-1)}}\right) \sum_{u=0}^{k-1} \chi^u p^{n(2k-u)(s-1)} p^{nu} + \chi^u p^{nks}. \tag{4.21}$$

It is not difficult to derive (4.21) directly.[5]

As for (4.01), it has now been proved that in *all* cases the number of solutions is determined by the same formula, namely (4.11). We state the

THEOREM 4.2. *If* $\alpha_1, \cdots, \alpha_{2s+1}$ *are non-zero elements of* $GF(p^n)$, *so numbered that*

$$(\alpha_1 + \alpha_2) \cdots (\alpha_{2s-1} + \alpha_{2s}) \neq 0,$$

and if in addition

$$\epsilon = \alpha_1 + \cdots + \alpha_{2s+1} \neq 0, \qquad \delta = \alpha_{2s+1},$$

[5] By a method similar to the proof of Theorem 2 of Representations.

the number of solutions of

$$\epsilon L = \alpha_1 X_1^2 + \cdots + \alpha_{2s+1} X_{2s+1}^2$$

in primary X_i *of degree* k *is given by* (4.11).

5. **Preliminary formulas for problem** (B). We now consider the number of sets of primary X_i of deg k, such that

$$\alpha L = \alpha_1 X_1^2 + \cdots + \alpha_{2s+1} X_{2s+1}^2, \qquad \alpha \neq 0, \tag{5.01}$$

where L is of degree $l < 2k$, and $\alpha_1 + \cdots + \alpha_{2s+1} = 0$. We assume the α_j so numbered that (3.01) holds; in the present case the possibility $\gamma = \alpha_1 + \cdots + \alpha_{2s} = 0$ is ruled out (for it would imply $\alpha_{2s+1} = 0$).

Writing $-\gamma$ for α_{2s+1}, it is evident that the number of solutions of (5.01) is

$$\sum_{\alpha L = \gamma F - \gamma X^2} n(\gamma F = \alpha_1 X_1^2 + \cdots + \alpha_{2s} X_{2s}^2), \tag{5.02}$$

the summation extending over all (primary) F, X of degrees $2k$, k, respectively, such that $\alpha L = \gamma F - \gamma X^2$. Employing (4.07), this becomes

$$\left(1 - \frac{\chi}{p^{n(s-1)}}\right) \sum_{\alpha L = \gamma F - \gamma X^2}^{m > k} \chi^m \mid M \mid^{s-1} + \sum_{\alpha L = \gamma F - \gamma X^2}^{m = k} \chi^m \mid M \mid^{s-1}$$

$$= \left(1 - \frac{\chi}{p^{n(s-1)}}\right) \sum_{U}^{u<k} \chi^u p^{n(2k-u)(s-1)} \sum_{\substack{\alpha L = \gamma MU - \gamma X^2 \\ U \text{ fixed}}} 1 + \sum_{\substack{\alpha L = \gamma MU - \gamma X^2 \\ \deg U = k}} \chi^k p^{nk(s-1)}$$

$$= \chi^k p^{nk(s-1)} W_k\left(-\frac{\alpha}{\gamma} L\right) \tag{5.03}$$

$$+ \left(1 - \frac{\chi}{p^{n(s-1)}}\right) \sum_{u=0}^{k-1} \chi^u p^{n(2k-u)(s-1)} p^{k-u} W_u\left(-\frac{\alpha}{\gamma} L\right),$$

by (2.10) and (4.12). Thus we have the following

THEOREM 5.1. *If* $\alpha\alpha_1 \cdots \alpha_{2s+1} \neq 0$, $\alpha_1 + \cdots + \alpha_{2s+1} = 0$, L *of degree* $l < 2k$, *the number of solutions of*

$$\alpha L = \alpha_1 X_1^2 + \cdots + \alpha_{2s+1} X_{2s+1}^2$$

in primary X_i *each of degree* k *is furnished by* (5.03).

6. **Transformation** $W_u(\Delta)$. Consider the product

$$p^{nk(s-1)} \sum_{u=0}^{\infty} \frac{W_u(\Delta)\chi^u}{p^{nuy}} \left\{1 + \left(1 - \frac{\chi}{p^{n(s-1)}}\right) \sum_{t=1}^{\infty} \frac{p^{nts}}{p^{nty}}\right\}$$

$$= p^{nk(s-1)} \sum_{u=0}^{\infty} \frac{W_u(\Delta)\chi^u}{p^{nuy}} \frac{1 - \chi p^{n(1-y)}}{1 - p^{n(s-y)}}, \tag{6.01}$$

for y sufficiently large. By properly choosing Δ, it is clear that the coefficient of p^{-nky} in the expansion of the left member of (6.01) may be identified with either (4.11) or (5.03).

We now transform

$$\sum_{u=0}^{\infty} \frac{W_u(\Delta)\chi^u}{p^{nuy}} = \sum_{U} \frac{\nu(\Delta, U)}{|U|^y} \chi^u \qquad (u = \deg U), \tag{6.02}$$

the summation in the right member extending over all primary U of degree $\geqq 0$. Using (2.02), the right member of (6.02) becomes

$$\prod_P \left\{1 + \frac{\nu(\Delta, P)}{|P|^y} \chi^\pi + \frac{\nu(\Delta, P^2)}{|P|^{2y}} \chi^{2\pi} + \cdots\right\} \qquad (\pi = \deg P), \tag{6.03}$$

the product extending over all primary irreducible P.

To evaluate (6.03) we make use of Lemmas 2.3 and 2.4. Take first $P \nmid \Delta$; then by (2.06), the corresponding term in (6.03) becomes

$$\frac{1 + \left(\dfrac{\Delta}{P}\right)\dfrac{\chi^\pi}{|P|^y}}{1 - \dfrac{\chi^\pi}{|P|^y}} = \frac{1 + \dfrac{\chi^\pi}{|P|^y}}{1 - \left(\dfrac{\Delta}{P}\right)\dfrac{\chi^\pi}{|P|^y}}. \tag{6.04}$$

For $P \mid \Delta$, it is convenient to consider two cases:

(i) $$P^{2\mu} \mid \Delta, \qquad P^{2\mu+1} \nmid \Delta.$$

In this case we use Lemma 2.4; a simple calculation shows that the corresponding term in the product (6.03) is

$$\frac{1 + \dfrac{\chi^\pi}{|P|^y}}{1 - \left(\dfrac{\Delta'}{P}\right)\dfrac{\chi^\pi}{|P|^y}}. \tag{6.05}$$

$$\left\{1 - \left(\frac{\Delta'}{P}\right)\frac{\chi^\pi}{|P|^y} + \frac{|P|}{|P|^{2y}} - \left(\frac{\Delta'}{P}\right)\frac{|P|\,\chi^\pi}{|P|^{3y}} + \frac{|P|^2}{|P|^{4y}} - \cdots + \frac{|P|^\mu}{|P|^{2\mu y}}\right\},$$

where $\Delta' = \Delta/P^{2\mu}$.

(ii) $$P^{2\mu+1} \mid \Delta,\ P^{2\mu+2} \nmid \Delta.$$

In this case it is easy to see that the corresponding term in (6.03) is

$$\left(1 + \frac{\chi^\pi}{|P|^y}\right)\left(1 + \frac{|P|}{|P|^{2y}} + \cdots + \frac{|P|^\mu}{|P|^{2\mu y}}\right), \tag{6.06}$$

which reduces to the first factor for $\mu = 0$.

Now from the definition of $\chi = \chi(\zeta)$ it is evident that $\chi^\tau = (\zeta/P)$. If then we put $\Delta = \Delta_1 F^2$, where F^2 is the largest primary square dividing Δ for which $(\Delta_1, F) = 1$, and substitute from (6.04), (6.05), (6.06) in (6.03), we have

$$\prod_{\text{all } P} \frac{1 + \frac{\chi^\tau}{|P|^y}}{1 - \left(\frac{\zeta\,\Delta_1}{P}\right)\frac{1}{|P|^y}} \cdot \tag{6.07}$$

$$\prod_{P\,|\,\Delta_1}\left(1 + \frac{|P|}{|P|^{2y}} + \cdots + \frac{|P|^\mu}{|P|^{2\mu y}}\right)\prod_{P\,|\,F} f(\zeta\Delta_1, P),$$

where we put (cf. second half of (6.05))

$$f(\Delta, P) = 1 - \left(\frac{\Delta}{P}\right)\frac{1}{|P|^y} + \frac{|P|}{P^{2y}} - \cdots + \frac{|P|^\mu}{P^{2\mu y}}. \tag{6.08}$$

This becomes somewhat simpler if we put $\Delta = \Delta_0 M^2$, where M^2 is the largest square dividing Δ, and therefore Δ_0 is quadratfrei. Formula (6.07) now becomes

$$\prod_{\text{all } P} \frac{1 + \frac{\chi}{|P|^y}}{1 - \left(\frac{\zeta\Delta_0}{P}\right)\frac{1}{|P|^y}} \prod_{P\,|\,M} f(\zeta\Delta_0, P), \tag{6.09}$$

where μ may be defined by $P^\mu \mid M$, $P^{\mu+1} \nmid M$.

As a further simplification, note that

$$\prod_P \left(1 + \frac{\chi^\tau}{|P|^y}\right) = \prod_P \frac{1 - |P|^{-2y}}{1 - \chi\,|P|^{-y}} = \frac{1 - p^{n(1-2y)}}{1 - \chi\, p^{n(1-y)}}.$$

Thus by (6.09) we see that the left member of (6.01) is

$$p^{nk(s-1)}\frac{1 - p^{n(1-2y)}}{1 - p^{n(s-y)}} \prod_{\text{all } P} \frac{1}{1 - \left(\frac{\zeta\,\Delta_0}{P}\right)\frac{1}{|P|^y}} \prod_{P\,|\,M} f(\zeta\,\Delta_0, P).$$

Recalling the remark at the beginning of this section, we have

Theorem 6.1. *If $L = L_0 M^2$, where L_0 is quadratfrei, and*

$$\vartheta = (-1)^s\,\epsilon\,\alpha_1\alpha_2 \cdots \alpha_{2s+1} \quad \text{or} \quad (-1)^s\,\alpha\,\alpha_1\alpha_2 \cdots \alpha_{2s+1}$$

according as $\deg L = 2k$ *or* $< 2k$, *the number of solutions of problem* (A) *or* (B), *respectively, is the coefficient of* p^{-nky} *in the product*

$$p^{nk(s-1)}\frac{1 - p^{n(1-2y)}}{1 - p^{n(s-y)}} \prod_P \left\{1 - \left(\frac{\vartheta L}{P}\right)\frac{1}{|P|^y}\right\}^{-1} \prod_{P\,|\,M} f(\vartheta L, M), \tag{6.10}$$

where $f(\vartheta L, P)$ *is defined by* (6.08).

7. Number of representations for quadratfrei L. If L is *quadratfrei*, the formulas of the preceding section are considerably shorter, and give rise to rather simple expressions for the number of representations. Thus (6.10) becomes

$$p^{nk(s-1)} \frac{1 - p^{n(1-2y)}}{1 - p^{n(s-y)}} \prod_P \left\{1 - \left(\frac{\vartheta L}{P}\right) \frac{1}{|P|^y}\right\}^{-1}. \tag{7.01}$$

Now let

$$\prod_P \left\{1 - \left(\frac{\vartheta L}{P}\right) \frac{1}{|P|^y}\right\}^{-1} = \sum_G \left(\frac{\vartheta L}{G}\right) \frac{1}{|G|^y},$$

the summation extending over all primary G. Grouping together polynomials of equal degree, the right member becomes

$$\sum_{j=0}^{\infty} \sigma_j\, p^{-njy}, \quad \text{where} \quad \sigma_j = \sum_{\deg G = j} \left(\frac{\vartheta L}{G}\right). \tag{7.02}$$

By the quadratic reciprocity theorem,[6]

$$\sigma_j = \pm \sum_{\deg G = j} \left(\frac{G}{L}\right) = 0 \quad \text{for} \quad j \geqq \deg L > 0. \tag{7.03}$$

The first sum in (7.02) therefore breaks off after a finite number of terms. Hence (7.01) becomes

$$p^{nk(s-1)} \frac{1 - p^{n(1-2y)}}{1 - p^{n(s-y)}} \sum_{j=0}^{l-1} \sigma_j\, p^{-njy}. \tag{7.04}$$

By Theorem 6.1 we are concerned only with the coefficient of p^{-nky}. Since

$$\frac{1 - p^{n(1-2y)}}{1 - p^{n(s-y)}} = 1 + \frac{p^{ns}}{p^{ny}} + \frac{p^{2ns} - p^n}{p^{2ny}} + \frac{p^{3ns} - p^{n(s+1)}}{p^{3ny}} + \cdots,$$

the coefficient of p^{-nky} in (7.04) is

$$p^{nk(s-1)} \{\sigma_k + p^{ns}\sigma_{k-1} + (p^{2ns} - p^n)\sigma_{k-2} + \cdots + (p^{nks} - p^{n(ks-2s+1)})\sigma_0\}. \tag{7.05}$$

This completes the proof of

THEOREM 7.1. *If* $\alpha_1, \cdots, \alpha_{2s+1}$ *are non-zero elements of* $GF(p^n)$ *such that*

$$\epsilon = \alpha_1 + \cdots + \alpha_{2s+1} \neq 0, \qquad \vartheta = (-1)^s \epsilon\, \alpha_1 \alpha_2 \cdots \alpha_{2s+1},$$

L primary quadratfrei of degree $2k$, the number of solutions of

$$\epsilon L = \alpha_1 X_1^2 + \cdots + \alpha_{2s+1} X_{2s+1}^2$$

in primary X_i of degree k is given by (7.05).

[6] R. Dedekind, Journal für Mathematik, vol. 54 (1857), pp. 1–26.

Similarly we have

THEOREM 7.2. *If $\alpha, \alpha_1, \cdots, \alpha_{2s+1}$ are non-zero elements of $GF(p^n)$ such that*

$$\epsilon = \alpha_1 + \cdots + \alpha_{2s+1} = 0, \qquad \vartheta = (-1)^s \alpha\alpha_1\alpha_2 \cdots \alpha_{2s+1},$$

L primary quadratfrei of degree $l < 2k$, the number of solutions of

$$\alpha L = \alpha_1 X_1^2 + \cdots + \alpha_{2s+1} X_{2s+1}^2 \tag{7.06}$$

in primary X_i of degree k is given by (7.05).

When $l = 0$, that is, $L = 1$, the formula (7.05) simplifies considerably. Indeed, if ϑ = square in $GF(p^n)$,

$$\prod_P \left\{1 - \left(\frac{\vartheta}{P}\right)\frac{1}{|P|^y}\right\}^{-1} = \prod_P \left(1 - \frac{1}{|P|^y}\right)^{-1} = \frac{1}{1 - p^{n(1-y)}},$$

so that in this case $\sigma_j = p^{nj}$. If $\vartheta \neq$ square,

$$\sigma_j = \sum_{\deg G = j} \left(\frac{\vartheta}{G}\right) = (-1)^j p^{nj}.$$

Hence, if we define $\chi = \chi(\vartheta)$ as in §4, namely $\chi = 1$ or -1 according as ϑ is or is not a square in $GF(p^n)$, (7.04) becomes

$$p^{nk(s-1)} \{\chi^k p^{nk} + \chi^{k-1} p^{n(s+k-1)} + \cdots + (p^{nks} - p^{n(ks-2s+1)})\}. \tag{7.07}$$

This expression is evidently the number of solutions of

$$\alpha = \alpha_1 X_1^2 + \cdots + \alpha_{2s+1} X_{2s+1}^2, \qquad \deg X_i = k.$$

Let us now suppose that $l > 0$. By (7.03), $\sigma_j = 0$ for $j \geqq l$. Thus when $l = 1$, $\sigma_0 = 1$, $\sigma_j = 0$ for $j \geqq 1$. Hence in this case (7.05) reduces to

$$p^{nk(s-1)} (p^{nks} - p^{n(ks-2s+1)}) \qquad (k \geqq 2),$$

which is therefore the number of solutions of (7.06) for L of degree 1. The number of solutions is evidently independent of the particular L appearing in the left member of (7.06).

When $l = 2$, σ_0 and σ_1 remain. The latter is determined by

$$\sigma_1 = \sum_\beta \left(\frac{\vartheta L}{x - \beta}\right) = \chi(\vartheta) \sum_\beta \left(\frac{L_\beta}{x - \beta}\right) = \chi(\vartheta) \sum_\beta \chi(L_\beta),$$

where L_β is the result of substituting β for x in L, and each summation extends over all β in $GF(p^n)$. Like results hold for larger values of l. Other simplifications may be effected by using certain formulas due to Artin. For example, if l is odd, $= 2m + 1$, say,

$$\sigma_{2m-i} = p^{n(m-i)} \sigma_i, \qquad \sigma_{2m} = p^{nm}.$$

However, we shall not take the space to elaborate this point, or to develop the connection with class number formulas.

Making use of (7.05), we can derive a recursion formula for N^k, the number of solutions of (7.06). If $k \geqq l + 1$, then (7.03) implies $\sigma_k = \sigma_{k-1} = 0$, so that (7.05) reduces to

$$p^{nk(s-1)}\,(p^{2ns} - p^n)\,(\sigma_{k-2} + p^{ns}\,\sigma_{k-3} + \cdots + p^{ns(k-2)}\,\sigma_0)\,.$$

Writing $k + 1$ for k, we have

$$N^{k+1} = p^{n(k+1)(s-1)}\,(p^{2ns} - p^n)\,(p^{ns}\,\sigma_{k-2} + p^{2ns}\,\sigma_{k-3} + \cdots + p^{ns(k-2)}\,\sigma_0)\,,$$

and therefore

$$N^{k+1} = p^{n(2s-1)}\,N^k\,.$$

Repeated application of this formula leads to

THEOREM 7.3. *If N^k denotes the number of solutions of* (7.06), *and L is of degree l, then*

$$N^{k+i} = p^{ni(2s-1)}\,N^k \qquad \textit{for} \qquad k > l > 0\,. \tag{7.08}$$

8. **Number of representations for general** L. Returning to the general formula (6.09), we put

$$L = L_0 M\,, \qquad \deg L = l\,, \tag{8.01}$$

$$\vartheta = (-1)^s\,\epsilon\alpha_1 \cdots \alpha_{2s+1} \qquad \text{or} \qquad (-1)^s\,\alpha\alpha_1 \cdots \alpha_{2s+1}$$

according as $l = 2k$ or $l < 2k$. L_0 is assumed quadratfrei. It will be convenient to break M into two factors, $M = M_1 M_2$, such that

$$M_1 = \prod_{P \mid L_0} P^\mu\,, \qquad M_2 = \prod_{P \nmid L_0} P^\mu\,; \tag{8.02}$$

in other words, M_2 is the greatest divisor of M that is prime to L_0. Then the second product in (6.09) gives rise to

$$\prod_{P \mid M_1}\left(1 + \frac{|P|}{|P|^{2y}} + \cdots + \frac{|P|^\mu}{|P|^{2\mu y}}\right) = \sum_{D_1 \mid M_1} \frac{|D_1|}{|D_1|^{2y}}\,;$$

and, for the factor M_2,

$$\prod_{P \mid M_2}\left\{1 - \left(\frac{\vartheta L_0}{P}\right)\frac{1}{|P|^y} + \cdots + \frac{|P|^\mu}{|P|^{2\mu y}}\right\} = \sum_{D_2 \mid M_2^2} \lambda(D_2)\left(\frac{\vartheta L_0}{D_2}\right)\frac{|D_2'|}{|D_2|^y}\,,$$

where $D_2'^{\,2}$ is the greatest square dividing D_2, and $\lambda(D_2)$ is the Liouville λ-function.[7]

Let us also put

$$\frac{1 - p^{n(1-2y)}}{1 - p^{n(s-y)}} \prod_P \left\{1 - \left(\frac{\vartheta L_0}{P}\right)\frac{1}{|P|^y}\right\}^{-1} = \sum_{j=0}^{\infty} N_j\,p^{-njy}\,, \tag{8.03}$$

[7] For M irreducible, $\lambda(M) = -1$; for $M = P_1 \cdots P_t$, $\lambda(M) = (-1)^t$.

so that

$$N_j = N_j(\vartheta L_0) = \sigma_j + p^{ns}\sigma_{j-1} + (p^{2ns} - p^n)\sigma_{j-2} + \cdots . \tag{8.04}$$

We wish to pick out the coefficient of p^{-nky} in the product

$$p^{nk(s-1)} \sum_j \frac{N_j}{p^{njy}} \sum_{D_1} \frac{|D_1|}{|D_1|^{2y}} \sum_{D_2} \lambda(D_2) \left(\frac{\vartheta L_0}{D_2}\right) \frac{|D_2'|}{|D_2|^y}.$$

If d_1, d_2, d_2' denote the degrees of D_1, D_2, D_2', respectively, the coefficient in question may be written in the form

$$p^{nk(s-1)} \sum_{D_1, D_2} \lambda(D_2) \left(\frac{\vartheta L_0}{D_2}\right) p^{n(d_1+d_2')} N_{k-2d_1-d_2}(\vartheta L_0), \tag{8.05}$$

the sum extending over all $D_1 \mid M_1$, $D_2 \mid M_2^2$; N_j defined by (8.04).

THEOREM 8.1. *The number of solutions X_i of*

$$\epsilon L = \alpha_1 X_1^2 + \cdots + \alpha_{2s+1} X_{2s+1}^2, \qquad \deg X_i = k, \qquad \deg L = 2k,$$

or of

$$\alpha L = \alpha_1 X_1^2 + \cdots + \alpha_{2s+1} X_{2s+1}^2, \qquad \deg X_i = k, \qquad \deg L < 2k,$$

is furnished by (8.05), *where* ϑ, L_0, M_1, M_2 *are determined by* (8.01) *and* (8.02).

The number of representations (8.05) simplifies in certain cases. If every irreducible divisor of M is a divisor of L_0, then $M_2 = 1$, $M = M_1$, and (8.05) becomes

$$p^{nk(s-1)} \sum_{D \mid M} p^{nd} N_{k-2d} \qquad (d = \deg D). \tag{8.06}$$

On the other hand, if $(L_0, M) = 1$, it follows that $M_1 = 1$, $M = M_2$; therefore (8.05) now reduces to

$$p^{nk(s-1)} \sum_{D \mid M^2} \lambda(D) \left(\frac{\vartheta L_0}{D}\right) p^{nd'} N_{k-d}, \tag{8.07}$$

where $d = \deg D$, $d' = \deg D'$, D'^2 the greatest square dividing D.

Another special case of some interest is $L_0 = 1$. Then it follows from (8.04) that

$$N_j = \chi^j p^{nj} + \chi^{j-1} p^{n(s+j-1)} + \chi^{j-2} p^{n(j-2)} (p^{2ns} - p^n) + \cdots . \tag{8.08}$$

In this case it is clear that $M_1 = 1$, $M = M_2$, so that (8.07) may be applied. Thus we get

$$p^{nk(s-1)} \sum_{D \mid M^2} \lambda(D) X^d p^{nd'} N_{k-d}. \tag{8.09}$$

THEOREM 8.2. *The number of solutions X_i of*

$$\epsilon M^2 = \alpha_1 X_1^2 + \cdots + \alpha_{2s+1} X_{2s+1}^2, \deg X_i = k = \deg M,$$

or of

$$\alpha M^2 = \alpha_1 X_1^2 + \cdots + \alpha_{2s+1} X_{2s+1}^2, \deg X_i = k > \deg M,$$

is furnished by (8.09), *where* N_j *is defined by* (8.08).

Making use of Theorem 7.3 it is possible to derive rather simple results concerning problem (B). Indeed, (7.08) implies

$$N_{j+i} = p^{nis} \quad \text{for} \quad j > l_0 = \deg L_0 > 0 . \tag{8.10}$$

This suggests that we choose L in such a manner that the subscript on N in (8.05) exceed l_0 for all D_1, D_2. The least value of the subscript is that corresponding to $D_1 = M_1$, $D_2 = M_2^2$, say $d_1 = m_1$, $d_2 = 2m_2$, so that we require $k - 2m_1 - 2m_2 > l_0$, that is, $k > l_0 + 2m_1 + 2m_2 = l = \deg L$. Then, for $k > \deg L$, (8.10) implies

$$N_{k-2d_1-d_2} = p^{ns(2m_1-2d_1+2m_2-d_2)} N_{k-2m_1-2m_2},$$

so that (8.05) becomes

$$p^{nk(s-1)} N_{k-2m} \sum_{D_1, D_2} \lambda(D_2) \left(\frac{\vartheta L_0}{D_2}\right) p^{ns(2m_1-2d_1+2m_2-d_2)} p^{n(d_1+d'_2)}$$

$$= p^{nk(s-1)} N_{k-2m} \cdot \sum_{M_1=D_1E_1} | D_1 E_1^{2s} | \cdot \sum_{M_2^2=D_2E_2} \lambda(D_2) \left(\frac{\vartheta L_0}{D_2}\right) | D'_2 E_2^s | \tag{8.11}$$

$$= p^{nk(s-1)} N_{k-2m} \cdot p^{nm_1} \zeta_{2s-1}(M_1) \cdot p^{nm_2} \zeta_{2s-1}(M_2, \vartheta L_0),$$

where

$$\zeta_{2s-1}(F) = \prod_{P \mid F} (1 + | P |^{2s-1} + \cdots + | P |^{\mu(2s-1)}) = \sum_{D \mid F} | D |^{2s-1},$$

$$\zeta_{2s-1}(F, \Delta) = \prod_{P \mid F} \left\{1 - \left(\frac{\Delta}{P}\right) | P |^{s-1} + | P |^{2s-1} - \left(\frac{\Delta}{P}\right) | P |^{3s-2} + \cdots + | P |^{\mu(2s-1)}\right\}, \tag{8.12}$$

and P^μ is the highest power of P dividing F.

But by (7.05)

$$p^{n(k-2m)} N_{k-2m} = p^{n(k-2m)} N_{k-2m}(\vartheta L_0)$$

is the number of solutions of

$$\alpha L_0 = \alpha_1 X_1^2 + \cdots + \alpha_{2s+1} X_{2s+1}^2, \qquad \deg X_i = k - 2m. \tag{8.13}$$

We may now state the

THEOREM 8.3. *If* L *is of degree* $l < k$, *and* $N^k(L)$ *denotes the number of solutions of*

$$\alpha L = \alpha_1 X_1^2 + \cdots + \alpha_{2s+1} X_{2s+1}^2, \qquad \deg X_i = k,$$

while $N^{k-2m}(L_0)$ *denotes the number of solutions of* (8.13), *then*

$$N^k(L) = p^{nm(2s-1)} N^{k-2m}(L_0) \zeta_{2s-1}(M_1) \zeta_{2s-1}(M_2, \vartheta L_0),$$

where L_0, M_1, M_2, ϑ, ζ_{2s-1} *are defined by* (8.01), (8.02), (8.13).

The following theorems are immediate corollaries:

THEOREM 8.4. *If for each irreducible* P, $P \mid L$, *the highest power of* P *dividing* L *is odd, then*

$$N^k(L) = p^{nm(2s-1)} N^{k-2m}(L_0) \zeta_{2s-1}(M).$$

For in this case it is evident that $M_1 = M$, $M_2 = 1$.

THEOREM 8.5. *If* $L = L_0M^2$, $(L_0, M) = 1$, L_0 *quadratfrei, then*

$$N^k(L) = p^{nm(2s-1)} N^{k-2m}(L_0) \zeta_{2s-1}(M, \vartheta L_0).$$

In particular, if M also is quadratfrei,

$$N^k(L) = p^{nm(2s-1)} N^{k-2m}(L_0) \prod_{P \mid M} \left\{1 - \left(\frac{\vartheta L_0}{P}\right) \mid P \mid^{s-1} + \mid P \mid^{2s-1}\right\}.$$

For in this instance $M_1 = 1$. The last result is the final one stated in the Introduction.

DUKE UNIVERSITY.

ASYMPTOTIC AND PRINCIPAL DIRECTIONS AT A PLANAR POINT OF A SURFACE

By Thomas L. Downs, Jr.

1. **Introduction.** If a regular point P of a real, analytic surface

$$S: \qquad x = x(u, v), \qquad y = y(u, v), \qquad z = z(u, v),$$

is not an umbilic, there exist at P two pairs of directions, the *asymptotic* and the *principal* directions, defined respectively by the equations

$$e\,du^2 + 2f du dv + g\,dv^2 = 0, \tag{1}$$

$$\begin{vmatrix} e\,du + f\,dv & E\,du + F\,dv \\ f\,du + g\,dv & F\,du + G\,dv \end{vmatrix} = 0, \tag{2}$$

where E, F, G and e, f, g, the coefficients in the first and second fundamental forms of S, are evaluated at P. These directions have the following well-known properties.

1. The tangent plane to S at P has contact of at least the second order in the asymptotic directions.

2. The normal curvature of S at P vanishes in the asymptotic directions and has its extrema in the principal directions.

3. The geodesic torsion of S at P vanishes in the principal directions.

4. The asymptotic lines are the solutions, for variable u and v, of the differential equation (1). Through each regular non-planar point there pass, in the asymptotic directions, two asymptotic lines.

5. The lines of curvature are the solutions of (2). Through each regular point which is not an umbilic there pass, in the principal directions, two real, distinct lines of curvature.

A *planar point* is defined as a regular point at which the quantities e, f, g vanish simultaneously. At a planar point the normal curvature and the geodesic torsion of the surface are zero in all directions, so that the first three theorems above become trivial. The fourth and fifth are false, because the point is singular for the differential equations (1) and (2).

In this paper there are defined two sets of directions at a planar point which play rôles similar to those played in the first three theorems by the asymptotic and principal directions at an ordinary point. These directions will be called the "true asymptotic" and the "true principal" directions. In a second paper the analogues of the fourth and fifth theorems will be established by studying the solutions at a planar point of the differential equations (1) and (2).

Received March 5, 1935; part of a thesis written at Harvard University under Professor W. C. Graustein.

2. The two sets of directions at a planar point. Let P be a real, regular point of the real, analytic surface S. By choosing P as the origin of the system of rectangular coördinates (x, y, z) and the tangent plane to S at P as the (x, y)-plane, we obtain as the representation of the surface in the neighborhood of P: $(0, 0, 0)$

$$z = F(x, y),$$

where $F(x, y)$ is an analytic function of (x, y) in a neighborhood of the point $(0, 0)$ and vanishes, together with its first partial derivatives, at this point:

$$F(0, 0) = 0, \qquad F_1(0, 0) = 0, \qquad F_2(0, 0) = 0.$$

If P is a planar point, that is, if

$$e = 0, \qquad f = 0, \qquad g = 0$$

at P, then

$$F_{11}(0, 0) = 0, \qquad F_{12}(0, 0) = 0, \qquad F_{22}(0, 0) = 0,$$

and Taylor's expansion of F about P becomes

$$z = \varphi_n(x, y) + \varphi_{n+1}(x, y) + \cdots, \qquad n > 2, \varphi_n(x, y) \not\equiv 0, \tag{3}$$

where φ_k is a homogeneous polynomial of degree k in x and y. The point P shall be called a *planar point of order n*.

It is to be noted that at P

$$E = 1, \qquad F = 0, \qquad G = 1,$$

and hence that the linear element of S at P, expressed in terms of $u = x$ and $v = y$ as curvilinear coördinates on S, is

$$ds^2 = du^2 + dv^2.$$

Order of contact of the tangent plane. We define the n directions determined at P by the equation

$$\varphi(du, dv) \equiv \varphi_n(du, dv) = 0 \tag{4}$$

as *the true asymptotic directions at P*. It follows from equation (3) that *the tangent plane at a planar point of order n has in general contact of order $(n - 1)$, but has contact of at least the n-th order in the true asymptotic directions and cuts the surface in those directions.* See Theorem 1, §1.

Normal curvature. Let a non-isotropic direction

$$du{:}dv = a{:}b, \qquad a^2 + b^2 = 1$$

be chosen arbitrarily at P, and let

$$C: \qquad u = as + \cdots, \qquad v = bs + \cdots$$

be any regular, analytic curve on S which passes through P in the chosen direction and is given in terms of its arc s. Then along C the normal curvature is

$$\frac{1}{\rho} = \frac{e\,du^2 + 2f\,dudv + g\,dv^2}{E\,du^2 + 2F\,dudv + G\,dv^2} = n(n-1)\,\varphi(a, b)\,s^{n-2} + \cdots .$$

Hence $1/\rho$ and its first $n-3$ derivatives with respect to s vanish at P regardless of the direction $du{:}dv$ chosen. But the quantity

$$\frac{1}{\sigma} = \frac{d^{n-2}}{ds^{n-2}}\left(\frac{1}{\rho}\right)_P = n!\,\frac{\varphi(du, dv)}{(du^2 + dv^2)^{n/2}} \tag{5}$$

varies with the direction at P and is independent of the curve C chosen in this direction. We shall call this quantity *the* $(n-2)$*-th directional derivative at* P *of the normal curvature.*

It follows from (5) that the $(n-2)$*-th directional derivative at* P *of the normal curvature vanishes in the true asymptotic directions, and in those directions only.* See Theorem 2, §1.

The analogy with Theorem 2 of the introduction can be carried farther. If we set in (5)

$$du = dr \cos\theta\,, \qquad dv = dr \sin\theta\,,$$

so that

$$\frac{1}{\sigma} = n!\,\varphi(\cos\theta, \sin\theta)\,, \tag{5'}$$

it is seen that *the possible extrema of* $1/\sigma$ *occur in the directions defined by the equation*

$$\psi(\cos\theta, \sin\theta) \equiv \frac{d}{d\theta}\,\varphi(\cos\theta, \sin\theta) = 0\,. \tag{6}$$

These directions we shall call *the true principal directions at* P.

We remark that the form $\psi(x, y)$ is one-half the jacobian of the form $\varphi(x, y)$ and the form

$$f(x, y) \equiv x^2 + y^2\,.$$

Geodesic torsion. Along the curve C the geodesic torsion of S is given by

$$\frac{1}{\tau} = \frac{\begin{vmatrix} e\,du + f\,dv & E\,du + F\,dv \\ f\,du + g\,dv & F\,du + G\,dv \end{vmatrix}}{D(E\,du^2 + 2F\,dudv + G\,dv^2)} = -\,(n-1)\,\psi(a, b)\,s^{n-2} + \cdots,$$

so that $1/\tau$ and its first $n-3$ derivatives with respect to s vanish at P, while the quantity

$$\frac{1}{\xi} = \frac{d^{n-2}}{ds^{n-2}}\left(\frac{1}{\tau}\right)_P = -\,(n-1)!\,\frac{\psi(du, dv)}{(du^2 + dv^2)^{n/2}} \tag{7}$$

varies with the direction at P and is independent of the particular curve C chosen in this direction. We shall call this quantity *the* $(n-2)$*-th directional derivative at* P *of the geodesic torsion.* It follows from (7) that *the* $(n-2)$*-th directional derivative at* P *of the geodesic torsion vanishes in the true principal directions and in those directions only.* See Theorem 3, §1.

Asymptotic lines and lines of curvature. Suppose that through the planar point P there passes an asymptotic line which is an analytic curve:

$$u = at^p + \cdots, \qquad v = bt^p + \cdots \qquad (p \geqq 1), \tag{8}$$

where a and b are not both zero. Since this curve is a solution of the differential equation of the asymptotic lines, the substitution of the series of (8) into equation (1) leads to the result that

$$\varphi(a, b) = 0,$$

and it follows that an analytic asymptotic line through a planar point is tangent there to a true asymptotic direction. More generally, it can be shown that *if an asymptotic line approaches a planar point in a definite direction, that direction is a true asymptotic direction.* The corresponding facts hold for the lines of curvature and the true principal directions. See Theorems 4 and 5, §1.

3. Projective and metric relations. The differentials (du, dv) are homogeneous projective coördinates in the pencil of tangent directions at P and have a metric interpretation in terms of the slope-angle θ, for the corresponding non-homogeneous coördinate is precisely the slope

$$dv/du = \tan\theta.$$

Thus, projective relations which exist among the roots of the three forms, φ, f, and their simultaneous covariant ψ, must also have an interpretation in terms of the angle θ. By a root of a form, say φ, we shall mean the slope corresponding to a solution (du, dv) of the homogeneous equation $\varphi(x, y) = 0$.

Let the initial direction for a slope as coördinate be chosen arbitarily, and let α_k, β_k be the respective cross-ratios of the roots a_k, b_k of φ, ψ with the two isotropic directions $f = 0$ and an arbitrarily chosen but fixed direction of slope λ:

$$\alpha_k = (i, -i; \lambda, a_k), \qquad \beta_k = (i, -i; \lambda, b_k).$$

Introduce new homogeneous projective coördinates (x', y') such that the isotropic directions assume the coördinates $(1, 0)$, $(0, 1)$ and the arbitrarily chosen direction of slope λ assumes the coördinate $(1, 1)$. Then the transform of f will be

$$f' \equiv \rho x'y', \qquad \rho \neq 0. \tag{9}$$

Hence, if the transform of φ is

$$\varphi' \equiv \sum_{k=0}^{n} c_k\, x'^k\, y'^{n-k}, \tag{10}$$

the transform of ψ will be

$$\psi' \equiv \sigma J(\varphi', f') \equiv \tau \sum_{k=0}^{n} (2k - n)c_k x'^k y'^{n-k}, \qquad \sigma\tau \neq 0. \tag{11}$$

Moreover, the roots of φ' and ψ':

$$y'/x' = a_k', \qquad y'/x' = b_k',$$

which are the new non-homogeneous coördinates of the true asymptotic and the true principal directions, respectively, will be precisely the cross-ratios α_k and β_k:

$$a_k' = (0\ \infty, 1a_k') = \alpha_k, \qquad b_k' = (0\ \infty, 1b_k') = \beta_k.$$

Suppose that $c_0c_n \neq 0$, that is, that none of the roots of φ' and ψ' is zero or infinite. Then none of the true asymptotic or principal directions coincides with an isotropic direction, and *the cross-ratios α_k, β_k will satisfy the n equations*

$$n\Sigma\beta_1\beta_2 \cdots \beta_k = (n - 2k)\Sigma\alpha_1\alpha_2 \cdots \alpha_k \qquad (k = 1, 2, \cdots, n). \tag{12}$$

In particular, if n is even,

$$\Sigma\beta_1\beta_2 \cdots \beta_m = 0 \qquad (n = 2m). \tag{13}$$

For comparison of the symmetric functions of the roots α_k of φ' with those of the roots β_k of ψ' yields the desired equations.

Inasmuch as S is a real surface and the planar point P is a real point on it, it follows that the two isotropic directions at P each count the same number of times as a true asymptotic (principal) direction. Moreover, it is clear from (9), (10), and (11) that *if the isotropic directions each count just p times as true asymptotic directions, they each count just p times as true principal directions. In this case, equations* (12) *and* (13) *hold for the cross-ratios*

$$\alpha_1, \alpha_2, \cdots, \alpha_{n-2p}; \qquad \beta_1, \beta_2, \cdots, \beta_{n-2p}$$

of the remaining true asymptotic and principal directions provided that in these equations we replace n by n − 2p.

It is to be noted that these results can be extended to yield a general theorem concerning the roots of the jacobian of an arbitrary binary form of the n-th degree and any binary quadratic form with distinct roots.

Metric interpretation. Let A_j, B_j be the directed angles which the true asymptotic directions and the true principal directions, respectively, make with the arbitrarily chosen direction of slope λ. From the Laguerre definition of angle we have that

$$\alpha_j = e^{2iA_j}, \qquad \beta_j = e^{2iB_j}.$$

Now the relations (12) are seen to be still true if α_j and β_j are replaced throughout by their reciprocals. Thus, the k-th equation of (12) yields the two relations

$$n\sum[\cos 2(B_1 + B_2 + \cdots + B_k) \pm i \sin 2(B_1 + B_2 + \cdots + B_k)]$$
$$= (n - 2k)\sum[\cos 2(A_1 + A_2 + \cdots + A_k) \pm i \sin 2(A_1 + A_2 + \cdots + A_k)].$$

Hence we obtain the $2n$ equations

$$n\sum \cos 2(B_1 + \cdots + B_k) = (n - 2k)\sum \cos 2(A_1 + \cdots + A_k),$$
$$(14) \quad n\sum \sin 2(B_1 + \cdots + B_k) = (n - 2k)\sum \sin 2(A_1 + \cdots + A_k)$$
$$(k = 1, 2, \cdots, n),$$

where the summation is to be taken over all terms of the form indicated. If n is even,

$$(15) \quad \sum \cos 2(B_1 + B_2 + \cdots + B_m) = \sum \sin 2 (B_1 + B_2 + \cdots + B_m) = 0$$
$$(2m = n).$$

The last pair of equations in (14) furnishes the interesting result: *The sum of the directed angles which the true asymptotic directions make with an arbitrarily chosen direction differs from the corresponding sum for the true principal directions by an odd number of right angles:*

$$(16) \quad (B_1 + B_2 + \cdots + B_n) = (A_1 + A_2 + \cdots + A_n) + \pi/2 \pm k\pi$$
$$(k = 0, 1, 2, \cdots).$$

Obvious modifications must be made in (14), (15), and (16) if the true asymptotic directions are not all distinct from the isotropic directions. These relations can be thrown into many different forms by trigonometric manipulation and the special choice of the direction from which the angles are measured.

Harmonic relations, $n = 4$. In this special case, it is possible to prove a number of projective theorems whose metric interpretation concerns the symmetrical arrangement of the true asymptotic and principal directions. It is to be noted that ψ is a form with real coefficients which always has at least one real root—see §4 below; hence if two true principal directions separate the isotropic directions harmonically, they must be real. It is also to be noted that the two isotropic directions are equally inclined to any non-isotropic direction. The theorems follow.

1. *Two true principal directions are perpendicular if and only if the four true principal directions form a harmonic set. In that case, the true asymptotic directions are symmetrically arranged with respect to the pair of perpendicular true principal directions.*

2. *The true principal directions consist of two pairs of perpendicular directions (which by Theorem* 1 *form a harmonic set) if and only if the true asymptotic directions are also perpendicular in pairs. In that case the true principal directions are the bisectors of the angles made by the pairs of non-perpendicular true asymptotic directions.*

3. *If the true asymptotic directions consist of two doubly-counting directions, each*

of these directions also counts once as a true principal direction and the angles made by them are bisected by the other two true principal directions.

4. *If a pair of true asymptotic directions bisect the angles made by the other two true asymptotic directions, they also bisect the angles made by each of two pairs of true principal directions.*

5. *If the four true asymptotic directions are symmetrically arranged* (*at angles of* 45°), *the true principal directions are similarly arranged and the directions of one kind bisect the angles made by adjacent directions of the other kind.*

As an example of the method by which these results are established, we shall outline the proof of the first theorem.

(i) If the isotropic directions each count once as true principal directions, the first statement in the theorem is trivial. Since the isotropic directions cannot each count twice as true principal directions, we may suppose in proving this statement that no true principal direction is an isotropic direction.

Then let there be given a projective change of coördinates which, as above, takes the isotropic directions $f = 0$ into the elements with the non-homogeneous coördinates 0 and ∞. Then equation (13) holds for the roots $\beta_1, \beta_2, \beta_3, \beta_4$ of ψ':

$$\beta_1\beta_2 + \beta_1\beta_3 + \beta_1\beta_4 + \beta_2\beta_3 + \beta_2\beta_4 + \beta_3\beta_4 = 0\,. \tag{13'}$$

The condition that the roots of ψ' form a harmonic set, $(\beta_1\beta_2, \beta_3\beta_4) = -1$, is, by direct computation,

$$2(\beta_1\beta_2 + \beta_3\beta_4) = (\beta_1 + \beta_2)(\beta_3 + \beta_4)\,. \tag{17}$$

But in view of (13′) this is equivalent to

$$(\beta_1 + \beta_2)(\beta_3 + \beta_4) = 0\,, \tag{18}$$

which is the condition, necessary and sufficient, that a pair of roots of ψ' separate harmonically the isotropic elements 0 and ∞.

(ii) Suppose now that the roots of ψ do form a harmonic set, and so contain a pair of real, perpendicular directions. Returning to the slope as coördinate, φ, f and ψ will have, to within constant multiples, the forms

$$\varphi = ax^4 + 4bx^3y + 6cx^2y^2 + 4dxy^3 + ey^4\,,$$

$$f = x^2 + y^2\,,$$

$$\psi = bx^4 + (3c - a)x^3y + 3(d - b)x^2y^2 + (e - 3c)xy^3 - dy^4\,.$$

Selecting the two perpendicular roots of ψ as the directions of slopes 0 and ∞, we have that $b = d = 0$, so that

$$\varphi = ax^4 + 6cx^2y^2 + ey^4\,,$$

whose roots are of the form $\lambda, -\lambda; \mu, -\mu$. The directions defined by these slopes, properly paired, separate harmonically the directions defined by the roots 0, ∞ of ψ. This completes the proof of the theorem.

4. **Qualitative relations.** We may also consider the directions under investigation as defined by the angles which satisfy the trigonometric equations

$$\varphi(\cos\theta, \sin\theta) = 0\,, \qquad \psi(\cos\theta, \sin\theta) \equiv \frac{d\varphi}{d\theta} = 0\,.$$

Then the following propositions are immediate consequences of the well-known relations between the zeros of a periodic function of class C' and the zeros of its derivative.

1. *Two real, distinct true asymptotic directions are always separated by a true principal direction. There is always at least one real true principal direction, and not all the true principal directions coincide with true asymptotic directions. There are at least as many real, distinct true principal directions as there are real, distinct true asymptotic directions.*

2. *If all the true asymptotic directions coincide with isotropic directions, every direction at P is a true principal direction.*

3. *If k true asymptotic directions coincide in a single direction, $k - 1$ true principal directions also coincide in that direction. Conversely, if m true principal directions coincide with an asymptotic direction, $m + 1$ true asymptotic directions coincide in that direction. If all n true asymptotic directions coincide in a single direction, $n - 1$ true principal directions coincide with that direction and the n-th is perpendicular to it.*

5. **Curvature of the surface near P.** In the representation

$$z = F(x, y) \equiv \varphi(x, y) + \cdots\,,$$

which we set up in §2 for S in the neighborhood of the planar point P: $(0, 0, 0)$, the coefficients of the second fundamental form have the values $e = F_{11}/D$, $f = F_{12}/D$, $g = F_{22}/D$, $D(0, 0) = 1$. The Gaussian curvature K of the surface is therefore given, near P, by the series

$$K = H(x, y) + \cdots\,,$$

where H is the Hessian of the form φ. Therefore H, provided it does not vanish identically, indicates the nature of K near P and the character at P of the curve of parabolic or planar points $K = 0$.

We list several well-known facts[1] concerning the Hessian $H_{2n-4}(x, y)$ of a form $\varphi_n(x, y)$. A root of φ is a multiple root of order p if and only if H has the same root to the order $(2p - 2)$. If $\varphi(a, b) = 0$, where a and b are real and not both zero, then $H(a, b) \leqq 0$; if (a, b) is a simple solution of $\varphi = 0$, then $H(a, b) < 0$. H vanishes identically if and only if $\varphi \equiv A(ax + by)^n$. Lastly, φ and H always vanish together if and only if

$$\varphi \equiv A(ax + by)^p(cx + dy)^q \qquad (p > 1, q > 1)\,.$$

[1] See, for example, F. Faa di Bruno: *Einleitung in die Theorie der binären Formen.*

By applying these facts to the case at hand, we establish the following results.

1. *The curve of planar or parabolic points* $K = 0$ *has a singular point of order at least* $(2n - 4)$ *at a planar point of order* n; *the order is greater than* $(2n - 4)$ *if and only if all the true asymptotic directions coincide.*

2. *A branch of the curve* $K = 0$ *is tangent at* P *to a true asymptotic direction if and only if two or more true asymptotic directions coincide in that direction. If the multiplicity of the true asymptotic direction is* p, *the multiplicity of the branch of the curve* $K = 0$ *which is tangent to it is* $(2p - 2)$.

3. *If* P *is an isolated planar point lying in the interior of a region in which* $K > 0$ *save for* P, *the true asymptotic directions are all imaginary and distinct; this is of course impossible if* n *is odd. On the other hand, if* P *is interior to a region in which* $K < 0$ *save for* P, *the true asymptotic directions are all distinct, but, for* $n > 3$, *not necessarily all real.*

4. *If each tangent to the curve* $K = 0$ *at* P *coincides with a true asymptotic direction and vice-versa, there are just two such directions, real or imaginary. If* $n - 1$ *true asymptotic directions coincide in a single, necessarily real, direction, all branches of the curve* $K = 0$ *at* P *are tangent to that direction.*

In the special case $n = 3$, sharper conclusions can be drawn. Let

$$\varphi(x, y) = ax^3 + 3bx^2y + 3cxy^2 + dy^3.$$

Without loss of generality, since n is odd, we can suppose that $a = 0$. The discriminant of the quadratic $H(1, t)$ is $\Delta = b^2(4bd - 3c^2)$, and that of the cubic $\varphi(1, t)$ is $-27\ \Delta$. The resultant of H and φ is Δ^2. An inspection of the two cases $\Delta > 0$, $\Delta < 0$ yields the following additional theorems, $n = 3$.

(i) *If there is but one real true asymptotic direction, the branches of the curve* $K = 0$ *are tangent at* P *to two real, distinct directions, neither of which is a true asymptotic direction.*

(ii) *There are three real, distinct true asymptotic directions if and only if* K *is negative in some deleted neighborhood of* P.

The case $\Delta = 0$ furnishes no information not given in propositions 1–4.

Locus of planar points. Finally, suppose that the planar point of order 3 lies on a curve of planar points. We may suppose that this curve is tangent at the origin P to the x-axis. Then, since e, f and g all vanish along the curve, we find that $\varphi(x, y) \equiv Ay^3$, $A \neq 0$; that is, *all three true asymptotic directions at* P *coincide in the direction of the planar locus.*

6. **Variation of the quantity** $1/\sigma$. **Indicatrix of Dupin.** At a non-planar point the indicatrix of Dupin is a conic, or a pair of conics, giving the variation of the normal curvature as a function of direction at the point and indicating the approximate shape of the surface near the point. At the planar point P, we have, in terms of the slope-angle, $1/\sigma = n!\,\varphi(\cos\theta, \sin\theta)$. Consider the point in the tangent plane for which $x = \sqrt[n]{|\sigma|}\cos\theta$, $y = \sqrt[n]{|\sigma|}\sin\theta$. As θ varies, this point describes the curve

$$n!\,\varphi(x, y) = \pm 1, \tag{19}$$

which consists of two "conjugate" n-tics, with asymptotes in the true asymptotic directions and "vertices" in the true principal directions. Since $1/\sigma$ changes sign only in the true asymptotic directions, only one of the two signs on the right of (19) is to be used in the region between any real pair of those directions.

The curve is evidently an *indicatrix* of the variation of $1/\sigma$ with θ. Moreover, it indicates the shape of the surface near P, for the section of S by a plane which is parallel to the tangent plane at P and near that plane will be a transcendental curve approximately similar to (19) or to certain branches of it.

Euler's Equation, $n = 3$. At a non-planar point, the normal curvature of S in the direction of slope-angle θ is given by

$$\frac{1}{\rho} = \frac{\cos^2\theta}{\rho_1} + \frac{\sin^2\theta}{\rho_2},$$

where $1/\rho_1$, $1/\rho_2$ are the principal normal curvatures. For a planar point of order 3 there is a generalized form of this equation, expressing the value of the quantity $1/\sigma$. It is

$$\frac{1}{\sigma} = \sum_{i,j,k}^{1,2,3} \frac{1}{\sigma_i} \frac{\sin(\theta - B_j)\sin(\theta - B_k)}{\sin(B_i - B_j)\sin(B_i - B_k)} \cos(\theta - B_i)\,, \tag{20}$$

where $i \neq j \neq k \neq i$ and i, j, k run cyclically, and where the B_i are the slope-angles of the true principal directions and the $1/\sigma_i$ are the values of $1/\sigma$ in those directions. The equation holds, of course, only if the slope-angles B_i are distinct. It has the advantage of being independent of the choice of the initial direction for the slope-angle, but it has the disadvantage of not characterizing the true principal directions. In fact, it holds for any three angles B_i and their corresponding values of $1/\sigma$ provided only that ΣB_i differs from the sum ΣA_i of the slope-angles of the true asymptotic directions by an odd multiple of $\pi/2$.

The deduction of equation (20) is direct. We have

$$1/\sigma = a\cos^3\theta + 3b\cos^2\theta\sin\theta + 3c\cos\theta\sin^2\theta + d\sin^3\theta\,. \tag{A}$$

By equation (16) of §3, the initial direction for the slope-angle can be so chosen that

$$A_1 + A_2 + A_3 = \pi\,, \qquad B_1 + B_2 + B_3 = \pi/2 \pm 2k\pi\,, \tag{B}$$

which imposes the condition $a = 3c$ upon the coefficients in (A). The values of $1/\sigma$ in the true principal directions are then given by the three equations

$$1/\sigma_i = a(\cos^3 B_i + \cos B_i \sin^2 B_i) + 3b\cos^2 B_i \sin B_i + d\sin^3 B_i \qquad (i = 1, 2, 3)\,. \tag{C}$$

These equations can be solved for the quantities a, b, c if the slope-angles B_i are distinct, for the determinant of the coefficients reduces to $\sin(B_2 - B_3)\sin(B_3 - B_1)\sin(B_1 - B_2) \neq 0$, provided we make use of the second relation of (B). The substitution of the solutions of (C) into (A) yields the desired equation (20).

7. Circular point. Suppose that P is a circular point that is, an umbilic which is not a planar point. Then at P: $e = E/\rho, f = F/\rho, g = G/\rho, 1/\rho \neq 0$. The differential equation of the lines of curvature has a singular point at P and the geodesic torsion vanishes in every direction there; therefore, we could define the quantity $1/\xi$ and the true principal directions $\psi = 0$ at P, as in §2. On the other hand, P is regular with respect to the asymptotic directions and the asymptotic lines through it; hence the quantity $1/\sigma$ is not defined. But, by means of a device suggested by a remark of Delloue's,[2] we can establish results which are formally similar to those of the preceding sections.

Let the axes be chosen as in §2. Then we shall have, in the neighborhood of P,

$$S: \qquad z = F(x, y)\,,$$

where now

$$F(0, 0) = 0\,, \qquad F_1(0, 0) = 0\,, \qquad F_2(0, 0) = 0\,,$$

$$F_{11}(0, 0) = e_0 = 1/\rho \neq 0\,, \qquad F_{12}(0, 0) = f_0 = 0\,, \qquad F_{22}(0, 0) = g_0 = 1/\rho\,.$$

The expansion of F about the origin P therefore becomes

$$z = (x^2 + y^2)/2\rho + \varphi_3(x, y) + \cdots + \varphi_k(x, y) + \cdots\,. \tag{21}$$

Consider the osculating sphere Σ at P; its equation can be written as

$$Z = (x^2 + y^2)/2\rho + f_3(x, y) + \cdots + f_k(x, y) + \cdots\,, \tag{22}$$

where

$$f_k \equiv 0, k \text{ odd}\,; \qquad f_k \equiv a_k(x^2 + y^2)^{k/2}/\rho^{k-1}, \qquad a_k \neq 0, \qquad k \text{ even}\,.$$

Instead of the distance from a point z on S to the tangent plane at P, we shall consider the distance from z to its vertical projection Z on Σ:

$$z - Z = \bar{\varphi}_n(x, y) + \bar{\varphi}_{n+1}(x, y) + \cdots \qquad (n > 2)\,.$$

In order that this expansion begin with terms of order n, it is necessary and sufficient that

$$\varphi_k(x, y) \equiv f_k(x, y) \qquad (k < n)\,.$$

The point P shall be called *a circular point of order n;* evidently, *the surface at a circular point of order n has in general contact of order $(n - 1)$ with its osculating sphere, but has contact of at least the n-th order in the directions given by*

$$\bar{\varphi}_n(du, dv) \equiv \varphi_n(du, dv) - f_n(du, dv) = 0\,.$$

The directions defined by $\bar{\varphi}_n = 0$ we may call "the true osculatory directions at P".

[2] Comptes Rendus, t. 187 (1928), p. 702. He also gives an interesting geometric interpretation of the true principal directions at a circular point.

We now note that the jacobians of the φ_k and $(x^2 + y^2)$,

$$\psi_k \equiv x \frac{\partial \varphi_k}{\partial y} - y \frac{\partial \varphi_k}{\partial x},$$

vanish identically for $k < n$, but that

$$\psi_n \equiv x \frac{\partial \varphi_n}{\partial y} - y \frac{\partial \varphi_n}{\partial x} \equiv x \frac{\partial \bar{\varphi}_n}{\partial y} - y \frac{\partial \bar{\varphi}_n}{\partial x}$$

does not in general vanish identically. Hence the methods of §2 lead us easily to the following conclusions. *A line of curvature which approaches a circular point in a definite direction must do so in a true principal direction. The quantity* $1/\xi$, defined as in equation (7), *vanishes in the true principal directions* $\psi_n = 0$ *and in those directions only.*

By analogy with §2, it is possible to define the quantity

$$\frac{1}{\bar{\sigma}} = \frac{n!\,\bar{\varphi}_n(du, dv)}{(du^2 + dv^2)^{n/2}},$$

which vanishes in the true osculatory directions at P and has its possible extrema in the true principal directions. It is the quantity $1/\sigma$ for the surface

$$\bar{S}: \qquad \bar{z} = z - Z,$$

which has a planar point of order n at P.

We see, then, that the theory of §§1–4, 6 which deals only with the directions defined by $\varphi_n = 0$, $\psi_n = 0$ at a planar point and with relations between these directions, is valid here when applied to the directions $\bar{\varphi}_n = 0$, $\psi_n = 0$ at a circular point. The developments of §5, concerning the curve $K = 0$ and the true asymptotic directions, can be applied here to the curve $\bar{K} = 0$ of $\bar{S}$ and the true osculatory directions. Unfortunately these developments do not, at any rate for $n > 3$, hold for the curve $K - 1/\rho^2 = 0$ of the given surface S.

SWEET BRIAR COLLEGE.

A MATHEMATICAL LOGIC WITHOUT VARIABLES. II

By J. B. Rosser

In the first part[1] we set up a formal system and proved that it was not too strong in the sense that it would not enable us to carry out certain (to us) undesirable types of proofs. We now prove that it is strong enough to do what we ask of it.

Section E

F1. P5 $\vdash EI$.

Proof.	$I = I \times I$	P5
	EI	M36

Strictly speaking, we should put

F1 $\to EI$
P5 $\vdash$ F1.

However, it is more convenient to write it as we have, thus stating on one line the fact that the formula which we are going to call F1 can be proved from P5. We shall in general omit explicit mention of M36, M37, or MC in the steps of a proof.

F2. P2 $\vdash EJ$.
F3. P1 $\vdash E\Pi$.
F4. P1 $\vdash E\Sigma$.
F5. P1 $\vdash ET$.
F6. P1 $\vdash EB$.
F7. P4 $\vdash EC$.
F8. P8 $\vdash EW$.
M38. $B(\mathbf{p} \times \mathbf{q})$ conv $B\mathbf{p} \times B\mathbf{q}$.

Proof.	$B(\mathbf{p} \times \mathbf{q})$ conv $(BB \times B)\mathbf{pq}$	r-conv.
	conv $(CB^3B)\mathbf{pq}$	P6
	conv $B\mathbf{p} \times B\mathbf{q}$	r-conv.

Note. $B(\mathbf{p} \times \mathbf{q})$ can be reduced to a form where no further reductions are possible. Therefore, by T14, if we perform all possible reductions on $(BB \times B)\mathbf{pq}$ we shall get the same result. By this means the reader can verify that $B(\mathbf{p} \times \mathbf{q})$ conv $(BB \times B)\mathbf{pq}$. Whenever this type of verification is possible, we omit the details of the proof. In the next step we indicate by the P6 to the right that P6 is the conversion postulate used in that step.

Received August 17, 1933 by the Editors of the Annals of Mathematics, accepted by them, and later transferred to this journal.

[1] The first part appeared in the Annals of Mathematics, vol. 36 (1935), pp. 127–150.

M39. $(\mathbf{p} \times \mathbf{q}) \times \mathbf{r}$ conv $\mathbf{p} \times (\mathbf{q} \times \mathbf{r})$

Proof. $(\mathbf{p} \times \mathbf{q}) \times \mathbf{r}$	conv $B(\mathbf{p} \times \mathbf{q})\mathbf{r}$	def. of $\times$
	conv $(B\mathbf{p} \times B\mathbf{q})\mathbf{r}$	M38
	conv $\mathbf{p} \times (\mathbf{q} \times \mathbf{r})$	r-conv.

F9. P5 $\vdash E1$.

Since BI is 1.

F10. P(1–16) $\vdash 1 = 1 \times 1$

Proof.	$1 = B(I \times I)$	F9, P5, P(1–16)
	$= 1 \times 1$	M38, P(1–16).

Throughout the rest of the paper, $\vdash \mathbf{q}$ shall be understood to mean P(1–16) $\vdash \mathbf{q}$.

F11. $\vdash E(B^2I)$. From P3.

F12. $\vdash B^2I = 1 \times B^2I$.

Proof. $B^2I = B(BI)$ (F11 and r-conv), $= (B^2I \times B)(BI)$(P3), $= 1 \times B^2I$ (r-conv).

F13. $\vdash B = 1 \times B$.

Proof. $B = B^2I \times B$ (P3), $= (1 \times B^2I) \times B$ (F12), $= 1 \times (B^2I \times B)$ (M39), $= 1 \times B$ (P3).

F14. $\vdash C = 1 \times C$. Similar proof.

F15. $\vdash W = 1 \times W$.

Proof. $W = W \times C$ (P15), $= W \times (B^2I \times C)$ (P4), $= (W \times B^2I) \times C$ (M39), $= (BW \times B^2)I \times C$ (r-conv), $= (CB^2W)I \times C$ (P8), $= (1 \times W) \times C$ (r-conv), $= 1 \times (W \times C)$ (M39), $= 1 \times W$ (P15).

Because of M39 we shall henceforth cease to distinguish between $(\mathbf{p} \times \mathbf{q}) \times \mathbf{r}$ and $\mathbf{p} \times (\mathbf{q} \times \mathbf{r})$, writing either as $\mathbf{p} \times \mathbf{q} \times \mathbf{r}$.

F16. $\vdash 1 = I \times 1 = 11 = 1^1$.

Note. This is a concise way of writing

$$\vdash 1 = I \times 1, \qquad \vdash I \times 1 = 11, \qquad \vdash 11 = 1^1.$$

Proof. $1 = BI$ (F9), $= (1 \times B)I$ (F13), $= 11$ (r-conv), $= I \times 1$ (def. of 1). By definition, 1^1 is 11.

F17. $\vdash B = I \times B = 1B = B^1$.

Proof. $B = 1 \times B$ (F13), $= (I \times 1) \times B$ (F16), $= I \times (1 \times B)$, $= I \times B$ (F13), $= 1B$.

F18. $\vdash C = I \times C = 1C = C^1$. Similar proof.

F19. $\vdash W = I \times W = 1W = W^1$. Similar proof.

M40. $I \times \mathbf{p}$ conv $\mathbf{p} \times I$.

Proof. $I \times \mathbf{p}$ r-conv $(BI \times I)\mathbf{p}$, conv $(CBI)\mathbf{p}$ (P9), r-conv $\mathbf{p} \times I$.

M41. $\mathbf{p}^{\mathbf{q}+\mathbf{r}}$ conv $\mathbf{p}^{\mathbf{q}} \times \mathbf{p}^{\mathbf{r}}$.

Proof. Use r-conversion and definitions of $\mathbf{p}^{\mathbf{q}}$ and $\mathbf{q} + \mathbf{r}$.

M42. $\mathbf{p}^{\mathbf{q}\times\mathbf{r}}$ conv $(\mathbf{p}^{\mathbf{r}})^{\mathbf{q}}$. Use r-conversion.

F20. $\vdash C = C \times B^2I = C \times 1$.

Proof. $C = B^2I \times C$ (P4), $= C^2 \times C$ (P12), $= (C \times C) \times C$ (r-conv and F18), $= C \times C^2$, $= C \times B^2I$ (P12), $= C \times B1$ (r-conv), $= C \times B(I \times 1)$ (F16), $= C \times B(1 \times I)$ (M40), $= C \times (B1 \times BI)$ (M38), $= (C \times B^2I) \times 1$, $= C \times 1$ (by the first part of proof).

F21. $\vdash W = W \times B^2I = W \times 1$.

Proof. $W = W \times C$ (P15), $= W \times C \times B^2I$ (F20), $= W \times B^2I$ (P15). $W = W \times 1$ by a similar argument.

F22. $\vdash 1 = CBI$.

Proof. $1 = I \times 1$ (F16), $= 1 \times I$ (M40), $= CBI$ (P9).

F23. $\vdash BC = BC \times B^3I$.

Proof. By P11, $\vdash E(BC)$, $\therefore BC = B(C \times B^2I)$ (F20), $= BC \times B(B^2I)$ (M38), $= BC \times B^3I$ (r-conv).

F24. $\vdash C \times B^3I = B^3I \times C$.

Proof. $\vdash I = (C \times B^3I)IIII$ (r-conv), $\therefore \vdash E(C \times B^3I)$, $\therefore C \times B^3I = (BC \times B^2)(BI)$ (r-conv), $= (CB^3C)(BI)$ (P7), $= B^3I \times C$ (r-conv).

F25. $\vdash B \times B^2I = B^3I \times B$.

Proof. $\vdash I = (B \times B^2I)IIII$ (r-conv), $\therefore \vdash E(B \times B^2I)$, $\therefore B \times B^2I = (BB \times B)(BI)$ (r-conv), $= (CB^3B)(BI)$ (P6), $= B^3I \times B$ (r-conv).

F26. $\vdash BB \times C = C \times BC \times B$.

Proof. $\vdash I = (BB \times C)IIII$ (r-conv), $\therefore \vdash E(BB \times C)$, $\therefore BB \times C = B(B^2I \times B) \times C$ (P3), $= B((1 \times B^2I) \times B) \times C$ (F12), $= B(1 \times B^2I) \times BB \times C$ (M38), $= B^2I \times B^3I \times BB \times C$ (M38 and r-conv), $= C^2 \times B^3I \times BB \times C$ (P12), $= C \times C \times B^3I \times BB \times C$ (r-conv and F18), $= C \times B^3I \times C \times BB \times C$ (F24), $= C \times B(C^2) \times C \times BB \times C$ (P12), $= C \times BC \times BC \times C \times BB \times C$ (M38, r-conv, and F18), $= C \times BC \times B \times C \times C$ (P10), $= C \times BC \times B \times B^2I$ (P12), $= C \times BC \times B^3I \times B$ (F25), $= C \times BC \times B$ (F23).

F27. $\vdash T = C1$.

Proof. $T = JII$, $= (W \times BC \times B^2C \times B^2B^2)III$ (P2), $= C(1 \times I)$ (r-conv), $= C(I \times 1)$ (M40), $= C1$ (F16).

Definition, by induction, of a *number*.

1. Any formula which is convertible into the formula 0 is a number.
2. If $\mathbf{n}$ is a number and $\mathbf{m}$ is convertible into the formula $\oplus\mathbf{n}$, then $\mathbf{m}$ is a number.

T19. 0 is a number.

T20. If $\mathbf{n}$ is a number, $\oplus\mathbf{n}$ is a number.

T21. $\oplus 0$ conv 1.

Proof. $\oplus 0$ r-conv 1(11), which by F16 and T18 is convertible into 1.

Corollary. 1 is a number.

Corollary. 2, 3, 4, etc., are numbers.

Because $\oplus 1$ conv 2, $\oplus 2$ conv 3, $\oplus 3$ conv 4, etc.

M43. If $\mathbf{n}$ occurs in $\mathbf{p}$ and there is an $\mathbf{f}$ such that:

1. $\mathbf{fn}$ conv $\mathbf{p}$.
2. $\vdash \mathbf{q}$ where $\mathbf{q}$ conv $\mathbf{f0}$.

3. From the assumption that **k** is a number and that **fk** conv **r** and **f**(⊕**k**) conv **s**, it follows that P(1–16), **r** $\vdash$ **s**.

then, if **n** is a number, $\vdash$ **p**.

Proof. We prove by induction that if the last two conditions of the theorem are satisfied, then if **n** is a number, $\vdash$ **fn**; $\vdash$ **p** then follows by the first condition.

1. If **k** conv 0, $\vdash$ **fk** by condition 2.

2. Given that **k** is a number and that $\vdash$ **fk**. Then $\vdash$ **r**, and by condition 3 $\vdash$ **s**, whence we get $\vdash$ **f**(⊕**k**). Then if **m** conv ⊕**k**, we get $\vdash$ **fm**.

M44. If **n** occurs in **p** and there is an **f** such that:

1. **fn** conv **p**.

2. $\vdash$ **q** where **q** conv **f**0.

3. $\vdash$ **r** where **r** conv **f**1.

4. From the assumption that **k** is a number and that **f**(⊕**k**) conv **s** and **f**($\oplus^2$**k**) conv **t**, it follows that P(1–16), **s** $\vdash$ **t**.

Then, if **n** is a number $\vdash$ **p**.

Proof. We show that conditions 3 and 4 imply condition 3 of M43. When **k** is a number, then either **k** conv 0, or there is an **m** such that **m** is a number and **k** conv ⊕**m**. Let $\mathbf{r}_1$ conv **fk**, and $\mathbf{s}_1$ conv **f**(⊕**k**).

1. **k** conv 0. Then ⊕**k** conv ⊕0, therefore ⊕**k** conv 1 by T21. Hence $\mathbf{s}_1$ conv **f**1 and so $\vdash \mathbf{s}_1$ by condition 3; hence P(1–16), $\mathbf{r}_1 \vdash \mathbf{s}_1$.

2. **m** is a number and **k** conv ⊕**m**. Then $\mathbf{r}_1$ conv **f**(⊕**m**) and $\mathbf{s}_1$ conv **f**($\oplus^2$**m**). Hence P(1–16), $\mathbf{r}_1 \vdash \mathbf{s}_1$ by condition 4.

F28. $\vdash 1 = \oplus 0$. Use F9 and T21.

F29. $\vdash 1 = 0\oplus$.

Proof. $1 = 1 \times 1$ (F10), $= 11 \times 11$ (F16), $= 0\oplus$ (r-conv).

T22. If **n** is a number, $\oplus^2$**n** conv S(⊕**n**).

Proof by induction.

1. If **k** conv 0, then $\oplus^2$**k** conv $\oplus^2 0$, conv $S1$ (r-conv), conv $S(\oplus 0)$ (T21), conv S(⊕**k**).

2. Assume that **k** is a number, that $\oplus^2$**k** conv S(⊕**k**), and that **m** conv ⊕**k**. Then $\oplus^2$**m** conv $\oplus^2$(⊕**k**), conv ⊕($\oplus^2$**k**) (r-conv), conv ⊕(S(⊕**k**)) (hyp. ind.), conv ⊕(S**m**), conv S(⊕**m**) (r-conv).

M45. If **n** is a number, $\vdash 1 = \mathbf{n}1 = 1^{\mathbf{n}}$.

Proof.

1. $C(B(\{ = \}1))$ 1**n** conv 1 = **n**1.

2. $\vdash 1 = 01$.

Because $1 = 11$ (F16), $= 01$ (r-conv).

3. $\vdash 1 = 11$ by F16.

4. If **n** is a number, then P(1–16), 1 = ⊕**n**1 $\vdash$ 1 = $\oplus^2$**n**1.

Because if 1 = ⊕**n**1, then 1 = 1 × ⊕**n**1 (F10), = 11 × ⊕**n**1 (F16), = S(⊕**n**)1 (r-conv), = $\oplus^2$**n**1 (T22).

Hence the theorem follows by M44.

COROLLARY. If **n** is a number, $\vdash E$**n**.

COROLLARY. If **n** is a number, $\vdash E$(⊕**n**) by T20.

COROLLARY. If $\mathbf{n}$ is a number, $\vdash E(\oplus^2\mathbf{n})$ since $\oplus^2\mathbf{n}$ conv $\oplus(\oplus\mathbf{n})$.

M46. If $\mathbf{n}$ is a number, $\vdash \oplus^2\mathbf{n} = S(\oplus\mathbf{n})$.

Proof. If $\mathbf{n}$ is a number, $\vdash E(\oplus^2\mathbf{n})$, $\therefore \oplus^2\mathbf{n} = S(\oplus\mathbf{n})$ by T22.

M47. If $\mathbf{n}$ is a number, then $\vdash \mathbf{n} = \mathbf{n}\oplus 0 = 0\oplus\mathbf{n} = \oplus^{\mathbf{n}}0 = \mathbf{n}^1$.

LEMMA 1. If $\mathbf{n}$ is a number, $\vdash \mathbf{n} = \mathbf{n}\oplus 0$.

Proof.

1. $(B^2W \times BC \times C \times B^2 \times C)\{ = \}\oplus 0\mathbf{n}$ conv $\mathbf{n} = \mathbf{n}\oplus 0$.

2. $\vdash 0 = 0\oplus 0$.

Because $0 = C(C11)1$ (M45 Cor.), $= (B^2I \times C)(C11)1$ (P4), $= 10$ (r-conv), $= 0\oplus 0$ (F29).

3. $\vdash 1 = 1\oplus 0$.

Because $1 = \oplus 0$ (F28), $= 1\oplus 0$ (r-conv).

4. If $\mathbf{k}$ is a number then P(1–16), $\oplus\mathbf{k} = (\oplus\mathbf{k})\oplus 0 \;\vdash \oplus^2\mathbf{k} = (\oplus^2\mathbf{k})\oplus 0$.

Because $\oplus^2\mathbf{k} = \oplus^2\mathbf{k}$ (M45 Cor.), $= \oplus(\oplus\mathbf{k})$ (r-conv), $= \oplus((\oplus\mathbf{k})\oplus 0)$ (hypothesis), $= S(\oplus\mathbf{k})\oplus 0$ (r-conv), $= (\oplus^2\mathbf{k})\oplus 0$ (M46).

Lemma is true by M44.

LEMMA 2. If $\mathbf{n}$ is a number, then $\mathbf{n} = 1\mathbf{n}$.

Proof.

1. $(BW \times B \times C)\{ = \}1\mathbf{n}$ conv $\mathbf{n} = 1\mathbf{n}$.

2. $\vdash 0 = 10$. See 2 of proof of Lemma 1.

3. $\vdash 1 = 11$ by F16.

4. If $\mathbf{k}$ is a number then P(1–16), $\oplus\mathbf{k} = 1(\oplus\mathbf{k}) \;\vdash \oplus^2\mathbf{k} = 1(\oplus^2\mathbf{k})$.

Because $\oplus^2\mathbf{k} = S(\oplus\mathbf{k})$ (M46), $= W(BC(B^2B(BB))1(\oplus\mathbf{k}))$ (r-conv), $= (1 \times W)(BC(B^2B(BB))1(\oplus\mathbf{k}))$ (F15), $= 1(S(\oplus\mathbf{k}))$ (r-conv), $= 1(\oplus^2\mathbf{k})$ (M46).

Lemma is true by M44.

LEMMA 3. If $\mathbf{n}$ is a number, $\vdash \mathbf{n} = 0\oplus\mathbf{n}$.

Proof. By Lemma 2, $\mathbf{n} = 1\mathbf{n}$, $= 0\oplus\mathbf{n}$ (F29).

Since $\mathbf{n}\oplus 0$ is $\oplus^{\mathbf{n}}0$ and $1\mathbf{n}$ is $\mathbf{n}^1$, the theorem is proved.

T23. If $\mathbf{n}$ and $\mathbf{m}$ are numbers, then $\mathbf{n}\oplus\mathbf{m}$ is a number.

Proof by induction on $\mathbf{n}$.

1. If $\mathbf{n}$ conv 0 and $\mathbf{m}$ is a number, then $\mathbf{n}\oplus\mathbf{m}$ is a number by M47 and T18.

2. Assume that $\mathbf{n}$ is a number and that if $\mathbf{m}$ is a number then $\mathbf{n}\oplus\mathbf{m}$ is a number. Also let $\mathbf{p}$ be conv $\oplus\mathbf{n}$. Either $\mathbf{n}$ conv 0, or there is a $\mathbf{k}$, which is a number, such that $\mathbf{n}$ conv $\oplus\mathbf{k}$.

Case 1. $\mathbf{n}$ conv 0. Then $\mathbf{p}$ conv 1 by T21. Hence $\mathbf{p}\oplus\mathbf{m}$ conv $1\oplus\mathbf{m}$ which is conv $\oplus\mathbf{m}$. Hence $\mathbf{p}\oplus\mathbf{m}$ is a number if $\mathbf{m}$ is a number.

Case 2. $\mathbf{k}$ is a number and $\mathbf{n}$ conv $\oplus\mathbf{k}$. Then $\mathbf{p}\oplus\mathbf{m}$ conv $\oplus^2\mathbf{k}\oplus\mathbf{m}$, conv $S(\oplus\mathbf{k})\oplus\mathbf{m}$ (T22), conv $S\mathbf{n}\oplus\mathbf{m}$, conv $\oplus(\mathbf{n}\oplus\mathbf{m})$. But, by the hypothesis of the induction, $\mathbf{n}\oplus\mathbf{m}$ is a number if $\mathbf{m}$ is a number and hence $\mathbf{p}\oplus\mathbf{m}$ is a number if $\mathbf{m}$ is a number.

COROLLARY. If $\mathbf{m}$ and $\mathbf{n}$ are numbers, $\vdash E(\mathbf{n}\oplus\mathbf{m})$.

M48. If $\mathbf{n}$ and $\mathbf{m}$ are numbers, $\vdash (1\oplus\mathbf{n})\oplus\mathbf{m} = 1\oplus(\mathbf{n}\oplus\mathbf{m})$.

Proof.

1. $\Phi\{ = \}(C(C(1\oplus)\oplus)\mathbf{m})(C(B(C(B(1\oplus))\mathbf{m}))\oplus)\mathbf{n}$ conv $(1\oplus\mathbf{n})\oplus\mathbf{m} = 1\oplus(\mathbf{n}\oplus\mathbf{m})$.

If $\mathbf{m}$ is a number:

2. $\vdash (1\oplus 0)\oplus\mathbf{m} = 1\oplus(0\oplus\mathbf{m})$ by M47.

3. If $\mathbf{k}$ is a number, then $\vdash (1\oplus(\oplus\mathbf{k}))\oplus\mathbf{m} = 1\oplus((\oplus\mathbf{k})\oplus\mathbf{m})$.

Because $(1\oplus(\oplus\mathbf{k}))\oplus\mathbf{m} = \oplus^2\mathbf{k}\oplus\mathbf{m}$ (r-conv and T23 Cor.), $= S(\oplus\mathbf{k})\oplus\mathbf{m}$ (M46), $= 1\oplus((\oplus\mathbf{k})\oplus\mathbf{m})$ (r-conv).

Theorem follows by M43.

Hereafter the use of T18 and T23 will be tacit.

M49. If $\mathbf{m}$, $\mathbf{n}$, and $\mathbf{p}$ are numbers, then $\vdash (\mathbf{m}\oplus\mathbf{n})\oplus\mathbf{p} = \mathbf{m}\oplus(\mathbf{n}\oplus\mathbf{p})$.

Proof.

1. $\Phi\{=\}((B^2C\times BC\times C\times T)\oplus\mathbf{n}\oplus\mathbf{p})((C\times T)\oplus(\mathbf{n}\oplus\mathbf{p}))\mathbf{m}$ conv $(\mathbf{m}\oplus\mathbf{n})\oplus\mathbf{p} = \mathbf{m}\oplus(\mathbf{n}\oplus\mathbf{p})$.

If $\mathbf{n}$ and $\mathbf{p}$ are numbers:

2. $\vdash (0\oplus\mathbf{n})\oplus\mathbf{p} = 0\oplus(\mathbf{n}\oplus\mathbf{p})$ by M47.

3. If $\mathbf{k}$ is a number and $(\mathbf{k}\oplus\mathbf{n})\oplus\mathbf{p} = \mathbf{k}\oplus(\mathbf{n}\oplus\mathbf{p})$, then $((\oplus\mathbf{k})\oplus\mathbf{n})\oplus\mathbf{p} = ((1\oplus\mathbf{k})\oplus\mathbf{n})\oplus\mathbf{p}$ (r-conv), $= (1\oplus(\mathbf{k}\oplus\mathbf{n}))\oplus\mathbf{p}$ (M48), $= 1\oplus((\mathbf{k}\oplus\mathbf{n})\oplus\mathbf{p})$ (M48), $= 1\oplus(\mathbf{k}\oplus(\mathbf{n}\oplus\mathbf{p}))$ (hypothesis), $= (1\oplus\mathbf{k})\oplus(\mathbf{n}\oplus\mathbf{p})$ (M48), $= (\oplus\mathbf{k})\oplus(\mathbf{n}\oplus\mathbf{p})$ (r-conv).

Theorem follows by M43.

M50. If $\mathbf{n}$ is a number, then $\vdash \mathbf{n}\oplus 1 = 1\oplus\mathbf{n}$.

Proof.

1. $\Phi\{=\}(C(C1\oplus)1)\oplus\mathbf{n}$ conv $\mathbf{n}\oplus 1 = 1\oplus\mathbf{n}$.

2. $\vdash 0\oplus 1 = 1\oplus 0$ by M47.

3. If $\mathbf{k}$ is a number and $\mathbf{k}\oplus 1 = 1\oplus\mathbf{k}$, then $(\oplus\mathbf{k})\oplus 1 = (1\oplus\mathbf{k})\oplus 1$ (r-conv), $= 1\oplus(\mathbf{k}\oplus 1)$ (M48), $= 1\oplus(1\oplus\mathbf{k})$ (hypothesis), $= 1\oplus(\oplus\mathbf{k})$ (r-conv).

Theorem follows by M43.

M51. If $\mathbf{m}$ and $\mathbf{n}$ are numbers, then $\vdash \mathbf{n}\oplus\mathbf{m} = \mathbf{m}\oplus\mathbf{n}$.

Proof.

1. $\Phi\{=\}(C(C1\oplus)\mathbf{m})(\mathbf{m}\oplus)\mathbf{n}$ conv $\mathbf{n}\oplus\mathbf{m} = \mathbf{m}\oplus\mathbf{n}$.

If $\mathbf{m}$ is a number:

2. $\vdash 0\oplus\mathbf{m} = \mathbf{m}\oplus 0$ by M47.

3. If $\mathbf{k}$ is a number and $\mathbf{k}\oplus\mathbf{m} = \mathbf{m}\oplus\mathbf{k}$, then $(\oplus\mathbf{k})\oplus\mathbf{m} = (1\oplus\mathbf{k})\oplus\mathbf{m}$ (r-conv), $= 1\oplus(\mathbf{k}\oplus\mathbf{m})$ (M48), $= 1\oplus(\mathbf{m}\oplus\mathbf{k})$ (hypothesis), $= (\mathbf{m}\oplus\mathbf{k})\oplus 1$ (M50), $= \mathbf{m}\oplus(\mathbf{k}\oplus 1)$ (M49), $= \mathbf{m}\oplus(1\oplus\mathbf{k})$ (M50), $= \mathbf{m}\oplus(\oplus\mathbf{k})$ (r-conv).

Theorem follows by M43.

The familiar properties which we have now proved for numbers and $\oplus$ will be used without reference in the future.

T24. If $\mathbf{m}$ and $\mathbf{n}$ are numbers, then $\mathbf{p}^{(\oplus\mathbf{m})\oplus(\oplus\mathbf{n})}$ conv $\mathbf{p}^{(\oplus\mathbf{m})+(\oplus\mathbf{n})}$.

Proof by induction on $\mathbf{m}$.

1. If $\mathbf{n}$ is a number and $\mathbf{m}$ conv 0, then $\mathbf{p}^{(\oplus\mathbf{m})\oplus(\oplus\mathbf{n})}$ conv $\mathbf{p}^{1\oplus(\oplus\mathbf{n})}$, conv $\mathbf{p}^{\oplus^2\mathbf{n}}$ (r-conv), conv $\mathbf{p}^{S(\oplus\mathbf{n})}$ (T22), conv $\mathbf{p}^{1+(\oplus\mathbf{n})}$ (def. of S and $+$), conv $\mathbf{p}^{(\oplus\mathbf{m})+(\oplus\mathbf{n})}$.

2. Assume that $\mathbf{m}$ is a number and that $\mathbf{p}^{(\oplus\mathbf{m})\oplus(\oplus\mathbf{n})}$ conv $\mathbf{p}^{(\oplus\mathbf{m})+(\oplus\mathbf{n})}$ if $\mathbf{n}$ is a number and that $\mathbf{k}$ conv $\oplus\mathbf{m}$. Then if $\mathbf{n}$ is a number $\mathbf{p}^{(\oplus\mathbf{k})\oplus(\oplus\mathbf{n})}$ conv $\mathbf{p}^{(1\oplus(1\oplus\mathbf{m}))\oplus(\oplus\mathbf{n})}$, conv $\mathbf{p}^{1\oplus(1\oplus(\mathbf{m}\oplus(\oplus\mathbf{n})))}$, conv $\mathbf{p}^{\oplus^2(\mathbf{m}\oplus(\oplus\mathbf{n}))}$ (r-conv), conv $\mathbf{p}^{S(\oplus(\mathbf{m}\oplus(\oplus\mathbf{n})))}$ (T22), conv $\mathbf{p}^{1+(1\oplus(\mathbf{m}\oplus(\oplus\mathbf{n})))}$ (r-conv), conv $\mathbf{p}^{1+((1\oplus\mathbf{m})\oplus(\oplus\mathbf{n}))}$, conv $\mathbf{p}^1\times\mathbf{p}^{(\oplus\mathbf{m})\oplus(\oplus\mathbf{n})}$ (r-conv),

conv $\mathbf{p}^1 \times \mathbf{p}^{(\oplus\mathbf{m})+(\oplus\mathbf{n})}$ (hyp. ind.), conv $\mathbf{p}^1 \times (\mathbf{p}^{\oplus\mathbf{m}} \times \mathbf{p}^{\oplus\mathbf{n}})$ (M41), conv $(\mathbf{p}^1 \times \mathbf{p}^{\oplus\mathbf{m}}) \times \mathbf{p}^{\oplus\mathbf{n}}$ (M39), conv $\mathbf{p}^{S(\oplus\mathbf{m})+(\oplus\mathbf{n})}$ (r-conv), conv $\mathbf{p}^{(\oplus\mathbf{k})+(\oplus\mathbf{n})}$ (T22 and r-conv).

T25. If $\mathbf{p}^0 \times \mathbf{p}^0$ conv $\mathbf{p}^0$, $\mathbf{p}^1 \times \mathbf{p}^0$ conv $\mathbf{p}^1$, $\mathbf{p}^0 \times \mathbf{p}^1$ conv $\mathbf{p}^1$, and $\mathbf{m}$ and $\mathbf{n}$ are numbers, then $\mathbf{p}^{\mathbf{m}\oplus\mathbf{n}}$ conv $\mathbf{p}^{\mathbf{m}+\mathbf{n}}$ and hence conv $\mathbf{p}^{\mathbf{m}} \times \mathbf{p}^{\mathbf{n}}$ by M41.

LEMMA 1. If $\mathbf{p}^1$ conv $\mathbf{p}^1 \times \mathbf{p}^0$ and $\mathbf{n}$ is a number, then $\mathbf{p}^{1\oplus\mathbf{n}}$ conv $\mathbf{p}^1 \times \mathbf{p}^{\mathbf{n}}$.

Proof. If $\mathbf{n}$ is a number, then either $\mathbf{n}$ conv 0 or there is a $\mathbf{k}$ such that $\mathbf{k}$ is a number and $\mathbf{n}$ conv $\oplus\mathbf{k}$. If $\mathbf{n}$ conv 0, then $\mathbf{p}^{1\oplus\mathbf{n}}$ conv $\mathbf{p}^1$ (M47), conv $\mathbf{p}^1 \times \mathbf{p}^0$ (hypothesis), conv $\mathbf{p}^1 \times \mathbf{p}^{\mathbf{n}}$. If $\mathbf{k}$ is a number and $\mathbf{n}$ conv $\oplus\mathbf{k}$, then $\mathbf{p}^{1\oplus\mathbf{n}}$ conv $\mathbf{p}^1 \times \mathbf{p}^{\mathbf{n}}$ by T21, T24 and M41.

LEMMA 2. If $\mathbf{p}^0 \times \mathbf{p}^0$ conv $\mathbf{p}^0$, $\mathbf{p}^1 \times \mathbf{p}^0$ conv $\mathbf{p}^1$, $\mathbf{p}^0 \times \mathbf{p}^1$ conv $\mathbf{p}^1$, and $\mathbf{n}$ is a number, then $\mathbf{p}^{0\oplus\mathbf{n}}$ conv $\mathbf{p}^{0+\mathbf{n}}$.

Proof. If $\mathbf{n}$ is a number, then either $\mathbf{n}$ conv 0 or there is a $\mathbf{k}$ such that $\mathbf{k}$ is a number and $\mathbf{n}$ conv $\oplus\mathbf{k}$. If $\mathbf{n}$ conv 0, then $\mathbf{p}^{0\oplus\mathbf{n}}$ conv $\mathbf{p}^0$, conv $\mathbf{p}^0 \times \mathbf{p}^0$ (hypothesis), conv $\mathbf{p}^{0+\mathbf{n}}$ (M41). If $\mathbf{k}$ is a number and $\mathbf{n}$ conv $\oplus\mathbf{k}$, then $\mathbf{p}^{0\oplus\mathbf{n}}$ conv $\mathbf{p}^{\oplus\mathbf{k}}$, conv $\mathbf{p}^{1\oplus\mathbf{k}}$, conv $\mathbf{p}^1 \times \mathbf{p}^{\mathbf{k}}$ (Lemma 1), conv $(\mathbf{p}^0 \times \mathbf{p}^1) \times \mathbf{p}^{\mathbf{k}}$ (hypothesis), conv $\mathbf{p}^0 \times (\mathbf{p}^1 \times \mathbf{p}^{\mathbf{k}})$, conv $\mathbf{p}^0 \times \mathbf{p}^{1\oplus\mathbf{k}}$ (Lemma 1), conv $\mathbf{p}^{0+\mathbf{n}}$ (M41).

We now prove the theorem by induction on $\mathbf{m}$.

1. If $\mathbf{n}$ is a number and $\mathbf{m}$ conv 0, then $\mathbf{p}^{\mathbf{m}\oplus\mathbf{n}}$ conv $\mathbf{p}^{\mathbf{m}+\mathbf{n}}$ by Lemma 2.

2. Assume that $\mathbf{m}$ is a number and that $\mathbf{p}^{\mathbf{m}\oplus\mathbf{n}}$ conv $\mathbf{p}^{\mathbf{m}+\mathbf{n}}$ if $\mathbf{n}$ is a number and that $\mathbf{k}$ conv $\oplus\mathbf{m}$. If $\mathbf{n}$ is a number, then $\mathbf{p}^{\mathbf{k}\oplus\mathbf{n}}$ conv $\mathbf{p}^{(1\oplus\mathbf{m})\oplus\mathbf{n}}$, conv $\mathbf{p}^{1\oplus(\mathbf{m}\oplus\mathbf{n})}$, conv $\mathbf{p}^1 \times \mathbf{p}^{(\mathbf{m}\oplus\mathbf{n})}$ (Lemma 1), conv $\mathbf{p}^1 \times (\mathbf{p}^{\mathbf{m}+\mathbf{n}})$ (hyp. ind.), conv $(\mathbf{p}^1 \times \mathbf{p}^{\mathbf{m}}) \times \mathbf{p}^{\mathbf{n}}$ (M41), conv $\mathbf{p}^{1\oplus\mathbf{m}} \times \mathbf{p}^{\mathbf{n}}$ (Lemma 1), conv $\mathbf{p}^{\mathbf{k}+\mathbf{n}}$ (M41).

F30. $\vdash 1 = B^0 = B^0 \times B^0$.

$\vdash B = B^1 = B^1 \times B^0 = B^0 \times B^1$.

Proof. $1 = 1 \times 1$ (F10), $= 0B$ (r-conv), $= B^0$(def.), $= B^0 \times B^0$ (F10). The rest of the theorem follows by F17, F13, and the following: $B = B^2I \times B$ (P3), $= CB^3BI$ (r-conv), $= (BB \times B)I$ (P6), $= B \times 1$ (r-conv).

T26. If $\mathbf{n}$ is a number, then $B(\mathbf{p}^{\oplus\mathbf{n}})$ conv $(B\mathbf{p})^{\oplus\mathbf{n}}$.

Proof by induction on $\mathbf{n}$.

1. When $\mathbf{n}$ conv 0, then $B(\mathbf{p}^{\oplus\mathbf{n}})$ conv $B(1\mathbf{p})$ (T21), conv $(B \times 1)\mathbf{p}$ (r-conv), conv $(1 \times B)\mathbf{p}$ (F30), conv $(B\mathbf{p})^1$ (r-conv), conv $(B\mathbf{p})^{\oplus\mathbf{n}}$ (T21).

2. Assume that $\mathbf{n}$ is a number and that $B(\mathbf{p}^{\oplus\mathbf{n}})$ conv $(B\mathbf{p})^{\oplus\mathbf{n}}$ and that $\mathbf{k}$ conv $\oplus\mathbf{n}$. Then $B(\mathbf{p}^{\oplus\mathbf{k}})$ conv $B(\mathbf{p}^{1\oplus(\oplus\mathbf{n})})$ (r-conv), conv $B(\mathbf{p}^1 \times \mathbf{p}^{\oplus\mathbf{n}})$ (T21 and T24), conv $B(\mathbf{p}^1) \times B(\mathbf{p}^{\oplus\mathbf{n}})$ (M38), conv $(B\mathbf{p})^1 \times (B\mathbf{p})^{\oplus\mathbf{n}}$ (hyp. ind. and (1) of this proof), conv $(B\mathbf{p})^{1\oplus(\oplus\mathbf{n})}$ (T21 and T24), conv $(B\mathbf{p})^{\oplus\mathbf{k}}$.

T27. If $\mathbf{n}$ is a number, then:

a. $B^{\mathbf{n}}B$ conv $B^{\mathbf{n}}B \times I$ conv $I \times B^{\mathbf{n}}B$.

b. $B^{\mathbf{n}}C$ conv $B^{\mathbf{n}}C \times I$ conv $I \times B^{\mathbf{n}}C$.

c. $B^{\mathbf{n}}W$ conv $B^{\mathbf{n}}W \times I$ conv $I \times B^{\mathbf{n}}W$.

d. $B^{\mathbf{n}}I$ conv $B^{\mathbf{n}}I \times I$ conv $I \times B^{\mathbf{n}}I$.

The proofs of these are just alike so we give only that of the second half of part (a) (since the first half follows from it by M40).

Proof. If $\mathbf{n}$ is a number, then either $\mathbf{n}$ conv 0 or there is a $\mathbf{k}$ such that $\mathbf{k}$ is a number and $\mathbf{n}$ conv $\oplus\mathbf{k}$. If $\mathbf{n}$ conv 0, then $B^{\mathbf{n}}B$ conv $1B$ (F30), conv $1(1B)$ (F17),

conv $I \times B^{\mathbf{n}}B$ (F30). If $\mathbf{k}$ is a number and $\mathbf{n}$ conv $\oplus\mathbf{k}$, then $B^{\mathbf{n}}B$ conv $B^{1\oplus\mathbf{k}}B$ (r-conv), conv $(B^1 \times B^{\mathbf{k}})B$ (T25 and F30), conv $B(B^{\mathbf{k}}B)$ (r-conv), conv $(1 \times B)(B^{\mathbf{k}}B)$ (F13), conv $I \times B(B^{\mathbf{k}}B)$ (r-conv), conv $I \times B^{\mathbf{n}}B$ (T25, F30, and r-conv).

It is obvious that there is an isomorphism between ordinary whole numbers and the combinations which we have called numbers. We set this up as follows:

1. Let 0 correspond to zero.
2. If $\mathbf{n}$ corresponds to n, let $\oplus\mathbf{n}$ correspond to n plus one.

By T24 and M41 it follows that if $\mathbf{n}$ corresponds to n and n is not zero then $\mathbf{p}^1 \times \mathbf{p}^1 \times \cdots \times \mathbf{p}^1$, where n factors occur, is convertible into $\mathbf{p}^{\mathbf{n}}$.

Moreover, if $\mathbf{n}$ is a number it follows by F30 and T25 that B^0 conv 1, B^1 conv B, and $B^{1\oplus\mathbf{n}}$ conv $B \times B^{\mathbf{n}}$ conv $BBB^{\mathbf{n}}$. Hence if $\mathbf{n}$ corresponds to n, $B^{\mathbf{n}}$ is convertible into the expression which Curry calls B_n except when n is zero (see Curry 1930, p. 529). Therefore we can carry most of Curry's proofs directly over into proofs in our system. However, in doing this we must note four things:

1. When a proof involves B_n, we must investigate the special case of n equal to zero.
2. Curry's "=" must be translated into "conv". Then if we can prove some function of the quantities concerned, we can use M13, M36, and MC.
3. Since five of Curry's axioms, namely $W \times K = BI$, $C \times K = BK$, $B \times K = K^2$, $CB^2K = BK \times I$, and $I = BI$, are not included in our list, we must make sure that these axioms are not used to prove the theorem in question.
4. Since K does not appear in our system, it must not appear in the formulas which we wish to carry over.

With this in mind we see that M42 is *Satz* 1, II B4, p. 532; M38 is *Satz* 2, II B4, p. 532; M39 is *Satz* 3, II B4, p. 533. *Satz* 4, II B4, p. 533 depends on the axiom $I = BI$ and so cannot be carried over. M40 is its nearest equivalent. *Satz* 5, II B4, p. 533 is a special case of T24.

T28. If $\mathbf{n}$ is a number, then $B^{\oplus\mathbf{n}}(\mathbf{p} \times \mathbf{q})$ conv $B^{\oplus\mathbf{n}}\,\mathbf{p} \times B^{\oplus\mathbf{n}}\,\mathbf{q}$.

Proof. *Satz* 6, II B4, p. 534.

T29. If $\mathbf{m}$, $\mathbf{n}$, and $\mathbf{p}$ are numbers, then $B^{\oplus\mathbf{m}}B^{\mathbf{n}} \times B^{\oplus\mathbf{m}}B^{\mathbf{p}} = B^{\oplus\mathbf{m}}B^{\mathbf{n}\oplus\mathbf{p}}$.

Proof. *Satz* 7, II B4, p. 534. Use F30 and T25 for the cases where $\mathbf{n}$ or $\mathbf{p}$ are conv 0.

From now on we shall essentially follow the treatment given in Curry 1930. However we shall use a slightly different definition of the normal form of a combinator, and actually show how to handle the permutators (see Curry 1930, p. 806) rather than referring them to group theory.

Definitions.[2]

$\Gamma \rightarrow (C \times T)(B^2CB)1.$

$\Delta \rightarrow (C \times T)(CB^2C)1.$

F31. $\vdash \Gamma\, 1 = C = \Delta 1.$

Proof. Use F20 and P4 because $\Gamma\, 1$ r-conv $C \times B^2I$ and $\Delta 1$ r-conv $B^2I \times C$.

T30. If $\mathbf{n}$ is a number, then

[2] Due to H. B. Curry.

a. $\Delta\ (\oplus^2\mathbf{n})$ conv $B(\Delta(\oplus\mathbf{n})) \times C$.

b. $\Gamma\ (\oplus^2\mathbf{n})$ conv $C \times B(\Gamma\ (\oplus\mathbf{n}))$.

Proof. Use T22 because $\Delta(S(\oplus\mathbf{n}))$ r-conv $B(\Delta(\oplus\mathbf{n})) \times C$ and $\Gamma\ (S(\oplus\mathbf{n}))$ r-conv $C \times B\ (\Gamma\ (\oplus\mathbf{n}))$.

T31. If $\mathbf{n}$ is a number, then

a. $B^{\oplus\mathbf{n}}C \times \Delta(\oplus\mathbf{n})$ conv $\Delta(\oplus^2\mathbf{n})$.

b. $\Gamma\ (\oplus\mathbf{n}) \times B^{\oplus\mathbf{n}}C$ conv $\Gamma\ (\oplus^2\mathbf{n})$.

Proof of (a) by induction on $\mathbf{n}$ (proof of (b) similar).

1. If $\mathbf{n}$ conv 0 use T30 and F31.

2. Assume the theorem for $\mathbf{k}$ and let $\mathbf{n}$ conv $\oplus\mathbf{k}$. Then $\Delta(\oplus^2\mathbf{n})$ conv $B(\Delta(\oplus\mathbf{n})) \times C$ (T30), conv $B(B^{\oplus\mathbf{k}}C \times \Delta(\oplus\mathbf{k})) \times C$ (hyp. ind.), conv $B(B^{\oplus\mathbf{k}}C) \times B(\Delta(\oplus\mathbf{k})) \times C$ (M38), conv $(B^1 \times B^{\oplus\mathbf{k}})C \times \Delta(\oplus^2\mathbf{k})$ (F30 and T30), conv $B^{1\oplus(\oplus\mathbf{k})}C \times \Delta(\oplus(\oplus\mathbf{k}))$ (T24), conv $B^{\oplus\mathbf{n}}C \times \Delta(\oplus\mathbf{n})$.

T32. If $\mathbf{m}$ and $\mathbf{n}$ are numbers, then $B^{\oplus\mathbf{m}}\mathbf{p}^{\oplus\mathbf{n}}$ conv $(B^{\oplus\mathbf{m}}\mathbf{p})^{\oplus\mathbf{n}}$.

Proof by induction on $\mathbf{m}$.

1. When $\mathbf{m}$ conv 0, use T26 and F30.

2. Assume the theorem for $\mathbf{k}$ and let $\mathbf{m}$ conv $\oplus\mathbf{k}$. If $\mathbf{n}$ is a number and $\mathbf{k}$ is a number, then $B^{\oplus\mathbf{m}}\mathbf{p}^{\oplus\mathbf{n}}$ conv $(B \times B^{\mathbf{m}})\mathbf{p}^{\oplus\mathbf{n}}$ (T25 and F30), conv $B(B^{\oplus\mathbf{k}}\mathbf{p}^{\oplus\mathbf{n}})$, conv $B((B^{\oplus\mathbf{k}}\mathbf{p})^{\oplus\mathbf{n}})$ (hyp. ind.), conv $(B(B^{\oplus\mathbf{k}}\mathbf{p}))^{\oplus\mathbf{n}}$ (T26), conv $(B^{\oplus\mathbf{m}}\mathbf{p})^{\oplus\mathbf{n}}$.

Section F

We shall follow closely Curry's treatment of combinators (see Curry 1930, pp. 799–834).

Definition. A combination is called a *combinator* if the only undefined terms appearing in it are J, I, (, and).

Where Curry uses the word "variable" we shall use instead "proper symbol." With this change, we take over the entire results of II C2 and II C3 (p. 799–803). The function which Curry calls W^n_{m+1} (p. 805) we will denote by $B^{\mathbf{m}}W^{\mathbf{n}}$ (where $\mathbf{n}$ and $\mathbf{m}$ correspond to n and m respectively) since they are convertible into each other by T24, M41, T32, F30, and F19. Making this change we take over *Satz* 3 and *Satz* 4, II C4, pp. 804–805. However, we shall make a different treatment of the permutation sequence (*Permutationsfolge*) from that of Curry.

We shall say that a combinator is a normal permutator if it is of the form $(\Delta 1)^{\mathbf{k}_1} \times (\Delta 2)^{\mathbf{k}_2} \times \cdots \times (\Delta\mathbf{n})^{\mathbf{k}_n}$ where $\mathbf{n}$, $\mathbf{k}_1, \cdots, \mathbf{k}_n$ are numbers and correspond respectively to n, $k_1, \cdots, k_n$ and $0 < k_i \leqq i + 1$ for $i < n$ and $0 < k_n \leqq n$.

T33. There is a unique normal permutator corresponding to each permutation sequence different from the identity.

Proof. We denote by σ_i the permutation sequence $(\mathbf{x}_0\mathbf{x}_{i+1}\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_i\mathbf{x}_{i+2} \cdots)$. Then by F31, T31, T24, and T27 $(\Delta\mathbf{i})^{\mathbf{k}}$ corresponds to σ_i^k, where σ_i^k stands for $\sigma_i \times \sigma_i \times \cdots \times \sigma_i$ where k factors occur (cf. *Satz* 1 and *Satz* 2, II C2, p. 800). Let the permutation sequence, π, be $(\mathbf{x}_0\mathbf{x}_{a_1}\mathbf{x}_{a_2} \cdots \mathbf{x}_{a_{n+1}}\mathbf{x}_{n+2} \cdots)$ where $a_{n+1} \neq n + 1$. n is at least one since the permutation sequence is not the

identity. We pick out the j (which will be unique since no repetitions or omissions occur in π) such that a_j is $n + 1$. Then, if π_1 is

$$(\mathbf{x}_0\mathbf{x}_{a_{j+1}}\,\mathbf{x}_{a_{j+2}} \cdots \mathbf{x}_{a_{n+1}}\mathbf{x}_{a_1}\mathbf{x}_{a_2} \cdots \mathbf{x}_{a_{j-1}}\mathbf{x}_{n+1}\mathbf{x}_{n+2} \cdots),\ \pi = \pi_1 \times \sigma_n^j\,.$$

We note that the $\mathbf{x}$'s have their proper order from $\mathbf{x}_{n+2}$ on in π whereas they have their proper order from $\mathbf{x}_{n+1}$ on in π_1. Now if π_1 is $(\mathbf{x}_0\mathbf{x}_{b_1} \cdots \mathbf{x}_{b_n}\,\mathbf{x}_{n+1} \cdots)$, we pick the k such that $b_k = n$ and find a π_2 such that $\pi_1 = \pi_2 \times \sigma_{n-1}^k$ and in π_2 the $\mathbf{x}$'s have their proper order from $\mathbf{x}_n$ on. Keeping this up we finally arrive at a unique sequence of k_i's so that π is $\sigma_1^{k_1} \times \cdots \sigma_n^{k_n}$. Then

$$(\Delta 1)^{\mathbf{k}_1} \times (\Delta 2)^{\mathbf{k}_2} \times \cdots \times (\Delta \mathbf{n})^{\mathbf{k}_n}$$

is a normal permutator and corresponds to π. By reversing the argument given above we readily show that it is the only normal permutator which corresponds to π.

As a matter of terminology, we shall call an "*Umwandlung*" which satisfies the conditions for the first factor in *Satz* 3, II C4, p. 804, a *duplication sequence*. We shall call a combinator defined as in *Satz* 4, II C4, p. 805, a *normal duplicator*. We shall call a "*Gruppierung*" defined as in *Festsetzung* 1, II C3, p. 801, a *composition sequence*. We shall call a combinator defined as in *Satz* 3, II C3, p. 802, a *normal compositor*. Finally, if $\mathbf{n}$ is a number, we shall call $B^{\mathbf{n}}I$ a *normal identity*. If a combinator has the form $\mathbf{X}_i$ $(0 < i < 5)$ or $\mathbf{X}_i \times \mathbf{X}_j$ $(0 < i < j < 5)$ or $\mathbf{X}_i \times \mathbf{X}_j \times \mathbf{X}_k$ $(0 < i < j < k < 5)$ or $\mathbf{X}_1 \times \mathbf{X}_2 \times \mathbf{X}_3 \times \mathbf{X}_4$ where $\mathbf{X}_1$ is a normal identity, $\mathbf{X}_2$ is a normal duplicator, $\mathbf{X}_3$ is a normal permutator, and $\mathbf{X}_4$ is a normal compositor, then we shall say that the combinator is *seminormal*.

By *Satz* 4, II C4, p. 805, a unique normal duplicator corresponds to each duplication sequence which is not the identity. We have shown that a unique normal permutator corresponds to each permutation sequence which is not the identity. By *Satz* 3, II C3, p. 802, a unique normal compositor corresponds to each composition sequence which is not the identity. Curry has shown[3] how to express any normal sequence without omissions in a unique manner as the product of a duplication sequence, a permutation sequence and a composition sequence and hence any combinator which corresponds to a normal sequence determines a unique permutation sequence. In particular (cf. *Satz* 1, II C2, p. 800) a seminormal combinator determines a unique permutation sequence.

Definition. If a seminormal combinator, $\mathfrak{R}$, has a normal identity and this normal identity is $B^{\mathbf{n}}I$ where $\mathbf{n}$ corresponds to n and $n + 1$ is the order with which $\mathfrak{R}$ corresponds to the normal sequence which $\mathfrak{R}$ determines (see Curry 1930, p. 790, *Festsetzung* 7), and if the unique permutation sequence which $\mathfrak{R}$ determines either

(a) is the identity and $\mathfrak{R}$ has no normal permutator; or

(b) is not the identity and the normal permutator of $\mathfrak{R}$ corresponds to it;

then $\mathfrak{R}$ is said to be a *normal* combinator.

[3] See pp. 553–554, Curry 1932.

Now (cf. *Satz* 2, II C5, p. 807) it is obvious that:

T34. If we have given a normal sequence η, and n is large enough so that a combinator can correspond to η with the order n, then there is a unique normal combinator which corresponds to η with order n.

By looking at Curry's definition of order (p. 790) we see that if a combinator, $\mathbf{p}$, corresponds with the order m to a sequence of variables, then $\mathbf{p}$ is of degree m by T8 and T9, and moreover $\mathbf{p}II \cdots I$, where I appears m times, is reducible to I and hence we can prove $\vdash E\mathbf{p}$.

M52. If $\mathbf{p}$ is a convertible into a normal combinator, then $\vdash E\mathbf{p}$.

Hereafter when $\mathbf{p}II \cdots I$ is reducible to I we shall not bother to write out the proof of $\vdash E\mathbf{p}$, even if $\mathbf{p}$ is not convertible into a normal combinator.

We continue with II D, p. 808, and modify Curry's definition of a regular combinator by saying that it must contain no K_n's.

Throughout the rest of Section F we shall let the following German and Greek letters stand for regular combinators of the form described after them.

$\mathfrak{B}$: only factors of the form $B^{\mathbf{m}}B^{\mathbf{n}}$.

$\mathfrak{C}$: only factors of the form $B^{\mathbf{m}}C$ or $\Gamma\,\mathbf{m}$ or $\Delta\,\mathbf{m}$.

$\mathfrak{J}$: only factors of the form $B^{\mathbf{n}}I$.

$\mathfrak{M}$: only factors of the form $B^{\mathbf{n}}W$ or $B^{\mathbf{n}}C$ or $\Gamma\,\mathbf{m}$ or $\Delta\mathbf{n}$.

$\mathfrak{N}$: a normal combinator.

$\mathfrak{W}$: only factors of the form $B^{\mathbf{n}}W$ or $B^{\mathbf{n}}W^{\mathbf{m}}$.

Ω: a combinator.

Satz 1, II D1, p. 809 is not valid in our system, because its proof depends on the axiom $I = BI$.

Satz 2, II D1, p. 809 is replaced by F26.

All of II D2 is valid[4] except *Satz* 5c and *Satz* 6. We wish to add to *Satz* 5 a part, d.

T35. If $\mathbf{m}$ and $\mathbf{n}$ are numbers, then $B^{\mathbf{m}}I \times B^{\mathbf{m}\oplus\mathbf{n}}\mathbf{p}$ conv $B^{\mathbf{m}\oplus\mathbf{n}}\mathbf{p} \times B^{\mathbf{m}}I$.

Proof. $CBI = 1$ (F22), $= 1 \times 1$ (F10), $= BI \times B^0$ (F30). Now apply *Satz* 3, II D2, p. 810.

We shall treat the identity combinator before proceeding with II D3. In order to save words we shall be a bit slack in our use of the correspondence between the combinations which are numbers and the ordinary whole numbers (as has already been done in the proof of T33). Thus if we say "$\mathbf{m}$ is greater than $\mathbf{n}$," it is understood that we mean "$\mathbf{m}$ corresponds to m, $\mathbf{n}$ corresponds to n, and m is greater than n."

M53. If we have a regular combinator $\mathfrak{J}$, and $B^{\mathbf{n}}$ is the highest power of B occurring in any of the factors $B^{\mathbf{m}}I$ of $\mathfrak{J}$, then $\vdash \mathfrak{J} = B^{\mathbf{n}}I$.

Proof. If $\mathbf{m}$ and $\mathbf{n}$ are numbers and $\mathbf{m} \leqq \mathbf{n}$, then $B^{\mathbf{n}}I \times B^{\mathbf{m}}I = B^{\mathbf{m}}I \times B^{\mathbf{n}}I = B^{\mathbf{n}}I$ by T28, T27d, and P5 and the theorem follows readily.

M54. If $\mathbf{m}$, $\mathbf{n}$, and $\mathbf{p}$ are numbers:

[4] Note the misprint in *Satz* 5a. It should read:

a) $$B_mY \cdot C_p = C_p \cdot B_mY, \qquad \text{wenn } m > p \geqq 1 \text{ gilt.}$$

a. If $\mathbf{m} \oplus 1 \geqq \mathbf{n}$, $\vdash B^{\mathbf{m}}B^{\mathbf{p}\oplus 1} \times B^{\mathbf{n}}I = B^{\mathbf{m}}B^{\mathbf{p}\oplus 1}$.

b. If $\mathbf{m} \oplus 1 \leqq \mathbf{n}$, $\vdash B^{\mathbf{m}}B^{\mathbf{p}\oplus 1} \times B^{\mathbf{n}}I = B^{\mathbf{n}\oplus\mathbf{p}\oplus 1}I \times B^{\mathbf{m}}B^{\mathbf{p}\oplus 1}$.

LEMMA 1. $B^{\mathbf{p}\oplus 1} = B^{\mathbf{p}\oplus 1} \times B^{1}I$. Since both equal $B^{\mathbf{p}\oplus 1} \times B^{0}$ by T25 and F30.

LEMMA 2. $B^{\mathbf{p}\oplus 1} = B^{\mathbf{p}\oplus 1} \times B^{0}I$. Since both equal $B^{\mathbf{p}\oplus 1} \times B^{1}I \times B^{0}I$ by Lemma 1 and M53.

LEMMA 3. $B^{\mathbf{p}\oplus 1} = B^{0}B^{\mathbf{p}\oplus 1}$. By Lemma 2, F30, P5, and M40.

LEMMA 4. If $\mathbf{q}$ is a number, $B^{\mathbf{q}}B^{\mathbf{p}\oplus 1} = B^{\mathbf{q}}B^{\mathbf{p}\oplus 1} \times B^{0}I$. Since both equal $B^{\mathbf{q}}B^{\mathbf{p}\oplus 1} \times B^{\mathbf{q}}I \times B^{0}I$ by T28 and Lemma 2 (or Lemma 2 and Lemma 3 if $\mathbf{q} = 0$) and M53.

Proof of theorem. If $\mathbf{m} \oplus 1 \leqq \mathbf{n}$, use *Satz* 4, II D2, p. 810. If $\mathbf{m} \oplus 1 = \mathbf{n}$, use T28 (or Lemma 3) and Lemma 1. If $\mathbf{m} \geqq \mathbf{n}$, use Lemma 4 and T28.

M55. If $\mathbf{m}$ and $\mathbf{n}$ are numbers:

a. If $\mathbf{m} \oplus 2 \geqq \mathbf{n}$, $\vdash B^{\mathbf{m}}C \times B^{\mathbf{n}}I = B^{\mathbf{m}}C$.

b. If $\mathbf{m} \oplus 2 \leqq \mathbf{n}$, $\vdash B^{\mathbf{m}}C \times B^{\mathbf{n}}I = B^{\mathbf{n}}I \times B^{\mathbf{m}}C$.

Proof. If $\mathbf{m} \oplus 2 \leqq \mathbf{n}$, use *Satz* 5a, II D2, p. 811. If $\mathbf{m} \oplus 2 = \mathbf{n}$ or $\mathbf{m} \oplus 1 = \mathbf{n}$, use F20 and T28. If $\mathbf{m} \geqq \mathbf{n}$, use T27b and T28.

M56. If $\mathbf{m}$, $\mathbf{n}$, and $\mathbf{p}$ are numbers:

a. If $\mathbf{m} \oplus 2 \geqq \mathbf{n}$, $\vdash B^{\mathbf{m}}W \times B^{\mathbf{n}}I = B^{\mathbf{m}}W$.

b. If $\mathbf{m} \oplus 1 \leqq \mathbf{p}$, $\vdash B^{\mathbf{m}}W \times B^{\mathbf{p}\oplus 1}I = B^{\mathbf{p}}I \times B^{\mathbf{m}}W$.

Proof. If $\mathbf{m} \oplus 1 \leqq \mathbf{p}$, use *Satz* 5b, II D2, p. 811. If $\mathbf{m} \oplus 2 = \mathbf{n}$ or $\mathbf{m} \oplus 1 = \mathbf{n}$, use F21 and T28. If $\mathbf{m} \geqq \mathbf{n}$, use T27c and T28.

T36. Any regular combinator can be converted into another regular combinator in which either no factor of the form $B^{\mathbf{n}}I$ ($\mathbf{n}$ a number) occurs or else only the first factor is of the form $B^{\mathbf{n}}I$.

Obvious by M53 to M56 inclusive.

All of II D3 is valid except *Satz* 5.

We now wish to show that if we have a regular combinator $\mathfrak{M}$, then there exists a normal identity $\mathfrak{J}$, a normal duplicator $\mathfrak{W}$ and a normal permutator $\mathfrak{C}$ such that $\mathfrak{J} \times \mathfrak{W} \times \mathfrak{C}$ is a normal combinator and $\vdash \mathfrak{M} = \mathfrak{J} \times \mathfrak{W} \times \mathfrak{C}$.[5] Hence we prove a series of theorems leading up to this.

M57. Any regular combinator which contains only terms of the form $B^{\mathbf{n}}W$, where $\mathbf{n}$ is a number, can be proved equal to a normal duplicator.

Proof. If $\mathbf{m}$ and $\mathbf{n}$ are numbers:

1. $\vdash B^{\mathbf{n}}W \times B^{\mathbf{m}\oplus 1}W = B^{\mathbf{m}}W \times B^{\mathbf{n}}W$ if $\mathbf{m} > \mathbf{n}$ by *Satz* 5b, II D2, p. 811.

2. $\vdash B^{\mathbf{n}}W \times B^{\mathbf{n}\oplus 1}W = B^{\mathbf{n}}W \times B^{\mathbf{n}}W = (B^{\mathbf{n}}W)^2 = B^{\mathbf{n}}W^2$ by P16, T28, T27c, T32 (or F30 and F19 if $\mathbf{n} = 0$).

By the definition of a normal duplicator and by T24, M41, and T32, the theorem is now obvious.

T37. Any regular combinator $\mathfrak{M}$ is convertible into a regular combinator of the form $\mathfrak{W} \times \mathfrak{C}$.

Proof. If $\mathbf{m}$ and $\mathbf{n}$ are numbers:

[5] This remark must be construed to include the case where one or both of the factors $\mathfrak{W}$ or $\mathfrak{C}$ is missing. A similar construction must be placed on T37.

1. If $\mathbf{m} > \mathbf{n}$, $\vdash B^{\mathbf{m}}C \times B^{\mathbf{n}}W = B^{\mathbf{n}}W \times B^{\mathbf{m}\oplus 1}C$ by *Satz* 5b, II D2, p. 811.

2. If $\mathbf{m} = \mathbf{n}$, $\vdash B^{\mathbf{m}}C \times B^{\mathbf{m}}W = B^{\mathbf{m}\oplus 1}W \times B^{\mathbf{m}}C \times B^{\mathbf{m}\oplus 1}C$ by P14 and T28.

3. If $\mathbf{m}\oplus 1 = \mathbf{n}$, $\vdash B^{\mathbf{m}}C \times B^{\mathbf{m}\oplus 1}W = B^{\mathbf{m}}W \times B^{\mathbf{m}\oplus 1}C \times B^{\mathbf{m}}C$.

$B^{\mathbf{m}}C \times B^{\mathbf{m}\oplus 1}W = B^{\mathbf{m}}C \times B^{\mathbf{m}\oplus 1}W \times B^{\mathbf{m}\oplus 2}I$	F21 and T28.
$= B^{\mathbf{m}}C \times B^{\mathbf{m}\oplus 1}W \times B^{\mathbf{m}}(C^2)$	P12, T25, F30.
$= B^{\mathbf{m}}C \times B^{\mathbf{m}\oplus 1}W \times B^{\mathbf{m}}C \times B^{\mathbf{m}}C$	F18, T28, *r*-conv, F30.
$= B^{\mathbf{m}}C \times B^{\mathbf{m}\oplus 1}W \times B^{\mathbf{m}\oplus 3}I \times B^{\mathbf{m}}C \times B^{\mathbf{m}}C$	F21 and T28.
$= B^{\mathbf{m}}C \times B^{\mathbf{m}\oplus 1}W \times B^{\mathbf{m}}C \times B^{\mathbf{m}\oplus 3}I \times B^{\mathbf{m}}C$	*Satz* 5a, II D2.
$= B^{\mathbf{m}}C \times B^{\mathbf{m}\oplus 1}W \times B^{\mathbf{m}}C \times B^{\mathbf{m}\oplus 1}(C^2) \times B^{\mathbf{m}}C$	P12, T24.
$= B^{\mathbf{m}}C \times B^{\mathbf{m}\oplus 1}W \times B^{\mathbf{m}}C \times B^{\mathbf{m}\oplus 1}C \times B^{\mathbf{m}\oplus 1}C \times B^{\mathbf{m}}C$	F18, T28
$= B^{\mathbf{m}}C \times B^{\mathbf{m}}C \times B^{\mathbf{m}}W \times B^{\mathbf{m}\oplus 1}C \times B^{\mathbf{m}}C$	part 2 of proof.
$= B^{\mathbf{m}}(C^2) \times B^{\mathbf{m}}W \times B^{\mathbf{m}\oplus 1}C \times B^{\mathbf{m}}C$	F18, T28, F30.
$= B^{\mathbf{m}\oplus 2}I \times B^{\mathbf{m}}W \times B^{\mathbf{m}\oplus 1}C \times B^{\mathbf{m}}C$	P12, T25, F30.
$= B^{\mathbf{m}}W \times B^{\mathbf{m}\oplus 3}I \times B^{\mathbf{m}\oplus 1}C \times B^{\mathbf{m}}C$	*Satz* 5b, II D2.
$= B^{\mathbf{m}}W \times B^{\mathbf{m}\oplus 1}C \times B^{\mathbf{m}}C$	P4, T28.

4. If $\mathbf{m}\oplus 1 < \mathbf{n}$, $\vdash B^{\mathbf{m}}C \times B^{\mathbf{n}}W = B^{\mathbf{n}}W \times B^{\mathbf{m}}C$ by *Satz* 5a, II D2, p. 811.

By use of 1, 2, 3 and 4, the theorem is now obvious because we identify C and B^0C because of F18 and F30.

Because of F31 and T31 it is obvious that we can speak of any $\mathfrak{C}$ as consisting only of factors of the form $B^{\mathbf{n}}C$. It is in this sense that Lemmas 15 and 16 of the following theorem are valid.

M58. If $\mathfrak{C}$ is a regular combinator, and $B^{\mathbf{n}}$ is the highest power of B occurring in any of the factors $B^{\mathbf{m}}C$ of $\mathfrak{C}$, then either there is a normal permutator, $\mathfrak{C}'$, of degree not more than $\mathbf{n}\oplus 3$ such that $\vdash \mathfrak{C} = B^{\mathbf{n}\oplus 2}I \times \mathfrak{C}'$ or $\vdash \mathfrak{C} = B^{\mathbf{n}\oplus 2}I$.

In order to simplify the notation we shall, if $\mathbf{n}$ is a number and corresponds to n, denote $B^{\mathbf{n}}C$ by C_{n+1}, $B^{\mathbf{n}}I$ by I_n, $\Gamma\,\mathbf{n}$ by Γ_n, and $\Delta\mathbf{n}$ by Δ_n. Further we shall make the temporary definition:

$$\mathfrak{S}_{n+1} \rightarrow \Gamma_n \times C_n \times \Delta_n .$$

Lemma 1. $\vdash C_n \times C_n = I_{n+1}$ by P12 and T28.

Lemma 2. $\vdash C_n = C_n \times I_m = I_m \times C_n$ if $m \leq n + 1$ by P4, F14, F20, T27b and T28.

Lemma 3. $\vdash C_{n+1} \times C_n \times C_{n+1} = C_n \times C_{n+1} \times C_n$ by P13 and T28.

Lemma 4. $\vdash C_n \times C_m = C_m \times C_n$ if $m > n + 1$ by *Satz* 5a, II D2, p. 811.

Lemma 5. $\vdash \Delta_n \times \Gamma_n = \Gamma_n \times \Delta_n = I_{n+1}$.

If we recall that Δ_n conv $C_n \times C_{n-1} \times \cdots \times C_2 \times C_1$ and Γ_n conv $C_1 \times C_2 \times \cdots \times C_{n-1} \times C_n$ then the lemma is obvious by Lemma 1, Lemma 2, and M53.

Lemma 6. If $m < n$, $\vdash C_m \times \Delta_n = \Delta_n \times C_{m+1}$.

$$C_m \times \Delta_n = C_m \times C_n \times C_{n-1} \times \cdots \times C_2 \times C_1$$

$$= C_n \times C_{n-1} \times \cdots \times C_{m+2} \times C_m \times C_{m+1} \times C_m \times C_{m-1} \times \cdots \times C_1 \qquad \text{Lemma 4.}$$

$$= C_n \times C_{n-1} \times \cdots \times C_{m+2} \times C_{m+1} \times C_m \times C_{m+1} \times C_{m-1} \times \cdots \times C_1 \qquad \text{Lemma 3.}$$

$$= \Delta_n \times C_{m+1} \qquad \text{Lemma 4.}$$

Lemma 7. $\vdash C_n \times \Delta_n = \Delta_n \times \mathfrak{S}_{n+1}$.

$$C_n \times \Delta_n = I_{n+1} \times C_n \times \Delta_n \qquad \text{Lemma 2.}$$

$$= \Delta_n \times \Gamma_n \times C_n \times \Delta_n \qquad \text{Lemma 5.}$$

$$= \Delta_n \times \mathfrak{S}_{n+1} \qquad \text{def.}$$

Lemma 8. $\vdash \mathfrak{S}_{n+1} \times \Delta_n = \Delta_n \times C_1$

$$\mathfrak{S}_{n+1} \times \Delta_n = \Gamma_n \times C_n \times \Delta_n \times \Delta_n$$

$$= \Gamma_n \times C_n \times \Delta_n \times C_n \times C_{n-1} \times \cdots \times C_2 \times C_1$$

$$= \Gamma_n \times C_n \times C_{n-1} \times C_{n-2} \times \cdots \times C_1 \times \Delta_n \times C_1 \qquad \text{Lemma 6.}$$

$$= I_{n+1} \times \Delta_n \times C_1 \qquad \text{Lemma 5.}$$

$$= \Delta_n \times C_1 \qquad \text{Lemma 2.}$$

Lemma 9. $\vdash \Delta_n = C_n \times C_{n-1} \times \cdots \times C_2 \times C_1$

$$= \mathfrak{S}_{n+1} \times C_n \times C_{n-1} \times \cdots \times C_3 \times C_2$$

$$= C_1 \times \mathfrak{S}_{n+1} \times C_n \times \cdots \times C_4 \times C_3$$

$$= C_2 \times C_1 \times \mathfrak{S}_{n+1} \times C_n \times \cdots \times C_5 \times C_4$$

. . . .

$$= C_{n-1} \times C_{n-2} \times \cdots \times C_2 \times C_1 \times \mathfrak{S}_{n+1}.$$

Proof. Take $p \leqq n$.

$$\Delta_n = I_{n+1} \times \Delta_n \qquad \text{Lemma 2.}$$

$$= \Gamma_n \times \Delta_n \times \Delta_n \qquad \text{Lemma 5.}$$

$$= \Gamma_n \times I_{n+1} \times \Delta_n \times \Delta_n \qquad \text{Lemma 2.}$$

$$= \Gamma_n \times \Gamma_n \times \Delta_n \times \Delta_n \times \Delta_n \qquad \text{Lemma 5.}$$

. . . .

$$= (\Gamma_n)^p \times \Delta_n \times (\Delta_n)^p \qquad \text{Lemmas 2 and 5.}$$

$$= (\Gamma_n)^p \times \Delta_n \times \mathfrak{S}_{n+1} \times C_n \times \cdots \times C_3 \times C_2 \times (\Delta_n)^{p-1} \qquad \text{Lemmas 6 and 7.}$$

$$= (\Gamma_n)^p \times \Delta_n \times \Delta_n \times C_1 \times \mathfrak{S}_{n+1} \times C_n \times \cdots \times C_3 \times (\Delta_n)^{p-2} \qquad \text{Lemmas 6, 7 and 8.}$$

. . . .

$$= (\Gamma_n)^p \times (\Delta_n)^p \times C_{p-1} \times C_{p-2} \times \cdots \times C_1 \times \mathfrak{S}_{n+1} \times \cdots \times C_{p+1} \qquad \text{Lemmas 6, 7, and 8.}$$

$$= I_{n+1} \times C_{p-1} \times C_{p-2} \times \cdots \times C_1 \times \mathfrak{S}_{n+1} \times \cdots \times C_{p+1} \qquad \text{Lemmas 2 and 5.}$$

$$= C_{p-1} \times C_{p-2} \times \cdots \times C_1 \times \mathfrak{S}_{n+1} \times \cdots \times C_{p+1} \qquad \text{M55 and Lemma 2.}$$

By taking values of p from 2 to n inclusive the Lemma is proved.

LEMMA 10. $\vdash (\Delta_n)^{\mathbf{n}\oplus 1} = I_{n+1}$ if $\mathbf{n}$ corresponds to n and $n > 0$.

Proof by induction.

1. $\vdash (C_1)^2 = I_2$ P12.
2. Assume the lemma true for $n = k$.

$$C_{k+1} \times C_{k+1} \times \Delta_k = I_{k+2} \times \Delta_k \qquad \text{Lemma 1.}$$

$$C_{k+1} \times \Delta_{k+1} = I_{k+2} \times \Delta_k \qquad \text{T31.}$$

$$C_{k+1} \times C_k \times C_{k-1} \times \cdots \times C_1 \times \mathfrak{S}_{k+2} = I_{k+2} \times \Delta_k \qquad \text{Lemma 9.}$$

However, by M55, I_{k+2} is commutative with Δ_k. Therefore

$$(I_{k+2} \times \Delta_k)^{\mathbf{k}\oplus 1} = (I_{k+2})^{\mathbf{k}\oplus 1} \times (\Delta_k)^{\mathbf{k}\oplus 1} \qquad \text{T24.}$$

$$= (I_{k+2})^{\mathbf{k}\oplus 1} \times I_{k+1} \qquad \text{hyp. ind.}$$

$$= I_{k+2} \qquad \text{M53 and T24.}$$

Hence $(C_{k+1} \times C_k \times \cdots \times C_1 \times \mathfrak{S}_{k+2})^{\mathbf{k}\oplus 1} = I_{k+2}$. Using T24 and regrouping we get

$$(C_{k+1} \times C_k \times \cdots \times C_1) \times (\mathfrak{S}_{k+2} \times C_{k+1} \times \cdots \times C_2) \times (C_1 \times \mathfrak{S}_{k+2} \times \cdots \times C_3) \times \cdots\cdots \times (C_k \times C_{k-1} \times \cdots \times C_1 \times \mathfrak{S}_{k+2}) = I_{k+2}.$$

But by Lemma 9, each factor is equal to Δ_{k+1} and there are $k + 2$ factors. Therefore $(\Delta_{k+1})^{\mathbf{k}\oplus 2} = I_{k+2}$ by T24.

LEMMA 11. If $\mathbf{p} < \mathbf{m} < \mathbf{n}\oplus 1$, $\vdash (\Delta_n)^{\mathbf{p}} \times C_m = C_{m-p} \times (\Delta_n)^{\mathbf{p}}$.

Obvious by Lemma 6 and T24.

LEMMA 12. If $\mathbf{m} < \mathbf{n}\oplus 1$,

$$\vdash (\Delta_n)^{\mathbf{m}} \times C_m = C_1 \times C_2 \times \cdots \times C_{n-1} \times (\Delta_n)^{\mathbf{m}\oplus 1}.$$

Proof. $(\Delta_n)^{\mathbf{m}} \times C_m = \Delta_n \times (\Delta_n)^{m-1} \times C_m$ T24.

$= \Delta_n \times C_1 \times (\Delta_n)^{m-1}$ Lemma 11.

$= \mathfrak{S}_{n+1} \times \Delta_n \times (\Delta_n)^{m-1}$ Lemma 8.

$= \Gamma_n \times C_n \times \Delta_n \times \Delta_n \times (\Delta_n)^{m-1}$ def.

$= C_1 \times C_2 \times \cdots \times C_{n-1} \times (\Delta_n)^{\mathbf{m}\oplus 1}$

Lemmas 1 and 2 and T24.

Lemma 13. If $\mathbf{m} < \mathbf{n}\oplus 1$,

$$\vdash (\Delta_n)^{\mathbf{m}\oplus 1} \times C_m = C_{n-1} \times C_{n-2} \times \cdots \times C_1 \times (\Delta_n)^{\mathbf{m}}$$

Proof. $(\Delta_n)^{\mathbf{m}\oplus 1} \times C_m = (\Delta_n)^2 \times C_1 \times (\Delta_n)^{m-1}$ T24 and Lemma 11.

$= \Delta_n \times \mathfrak{S}_{n+1} \times (\Delta_n)^{\mathbf{m}}$ T24 and Lemma 8.

$= \Delta_n \times \Gamma_n \times C_n \times \Delta_n \times (\Delta_n)^{\mathbf{m}}$ def.

$= C_{n-1} \times C_{n-2} \times \cdots \times C_1 \times (\Delta_n)^{\mathbf{m}}$ M55 and Lemmas 1, 2, and 5.

Lemma 14. If $\mathbf{m}\oplus 1 < \mathbf{p} < \mathbf{n}\oplus 1$, $\vdash (\Delta_n)^{\mathbf{p}} \times C_m = C_{m+n+1-p} \times (\Delta_n)^{\mathbf{p}}$.

Proof. Put $\mathbf{p} = \mathbf{m}\oplus 1 \oplus \mathbf{h}$.

$(\Delta_n)^{\mathbf{p}} \times C_m = (\Delta_n)^{\mathbf{h}\oplus 2} \times C_1 \times (\Delta_n)^{m-1}$ Lemma 11 and T24.

$= (\Delta_n)^{\mathbf{h}\oplus 1} \times \mathfrak{S}_{n+1} \times (\Delta_n)^{\mathbf{m}}$ Lemma 8 and T24.

$= (\Delta_n)^{\mathbf{h}} \times C_n \times (\Delta_n)^{\mathbf{m}\oplus 1}$ Lemma 7 and T24.

$= C_{n-h} \times (\Delta_n)^{\mathbf{h}\oplus\mathbf{m}\oplus 1}$ Lemma 11 and T24.

$= C_{n+m+1-p} \times (\Delta_n)^{\mathbf{p}}$

Lemma 15. If we have $\mathfrak{C}$ and no higher power of B than $B^{\mathbf{n}}$ occurs in any of the factors $B^{\mathbf{m}}C$ of $\mathfrak{C}$, then for every $\mathbf{r}$ there is an $\mathbf{s}$ and a $\mathfrak{C}'$ such that:

1. No higher power of B than $B^{\mathbf{n}}$ occurs in any of the factors $B^{\mathbf{m}}C$ of $\mathfrak{C}'$.
2. $\vdash (\Delta(\mathbf{n}\oplus 1))^{\mathbf{r}} \times \mathfrak{C} = \mathfrak{C}' \times (\Delta(\mathbf{n}\oplus 1))^{\mathbf{s}}$.

Obvious from Lemmas 10–14.

Lemma 16. If we have $\mathfrak{C}$ and the highest power of B occurring in any of the factors $B^{\mathbf{m}}C$ of $\mathfrak{C}$ is $B^{\mathbf{n}\oplus 1}$, then there is an $\mathbf{s}$ and a $\mathfrak{C}'$ such that:

1. No higher power of B than $B^{\mathbf{n}}$ occurs in any of the factors $B^{\mathbf{m}}C$ of $\mathfrak{C}'$.
2. $\mathbf{s} \leqq \mathbf{n}\oplus 3$.
3. $\vdash \mathfrak{C} = B^{\mathbf{n}\oplus 3}I \times \mathfrak{C}' \times (\Delta(\mathbf{n}\oplus 2))^{\mathbf{s}}$.

In case $\mathbf{s} = \mathbf{n}\oplus 3$, it follows from Lemma 10, and M55 that $\vdash \mathfrak{C} = B^{\mathbf{n}\oplus 3}I \times \mathfrak{C}'$.

Proof. Pick out the factors $B^{\mathbf{m}}C$ of $\mathfrak{C}$ for which $\vdash \mathbf{m} = \mathbf{n}\oplus 1$. Then by Lemmas 2 and 5, $\vdash B^{\mathbf{n}\oplus 1}C \times \Delta(\mathbf{n}\oplus 1) \times \Gamma(\mathbf{n}\oplus 1) = B^{\mathbf{n}\oplus 1}C$, $\therefore \vdash \Delta(\mathbf{n}\oplus 2) \times \Gamma(\mathbf{n}\oplus 1) = B^{\mathbf{n}\oplus 1}C$.

We replace each occurrence of $B^{\mathbf{n}\oplus 1}C$ by its equal. However $\Delta(\mathbf{n}\oplus 2) = B^{\mathbf{n}\oplus 3}I \times \Delta\,(\mathbf{n}\oplus 2)$, so we can get $B^{\mathbf{n}\oplus 3}I \times \mathfrak{C}'$. Then by Lemma 15 we move

all the $\Delta(\mathbf{n}\oplus 2)$'s over to the right-hand side, getting $B^{\mathbf{n}\oplus 3}I \times \mathfrak{C}'' \times (\Delta(\mathbf{n}\oplus 2))^{\mathbf{k}}$. Finally, because of Lemma 10, we can reduce $\mathbf{k}$ modulo $\mathbf{n}\oplus 3$.

We note that the degree of $\Delta(\mathbf{n}\oplus 2)$ is $\mathbf{n}\oplus 4$.

We now prove Theorem M58 by induction, using P12 and Lemma 16.

M59. Two regular combinators of the form $\mathfrak{C}_1$ and $\mathfrak{C}_2$ can be proved equal if, and only if, they correspond with the same order to the same permutation sequence.

By T16, T17, and T18, it follows that two combinators which correspond to permutation sequences correspond with the same order to the same sequence if they can be proved equal.

Assume that both $\mathfrak{C}_1$ and $\mathfrak{C}_2$ correspond with the order n to the permutation sequence π. Then each has the degree n. If either of them equals a normal identity (see M58), then π is the identity sequence and each must equal a normal identity by M58, since no normal permutator can correspond to the identity sequence. The two normal identities must then be equal since both are of degree n. In the case that neither $\mathfrak{C}_1$ or $\mathfrak{C}_2$ is equal to a normal identity, then by M58 if $B^{\mathbf{k}}$ is the highest power of B that occurs in any of the factors $B^{\mathbf{m}}C$ of $\mathfrak{C}_1$ and $B^{\mathbf{l}}$ is the highest power of B that occurs in any of the factors $B^{\mathbf{m}}C$ of $\mathfrak{C}_2$ there is a normal permutator $\mathfrak{C}_1'$ of degree $\mathbf{k}\oplus 3$ or less and a normal permutator $\mathfrak{C}_2'$ of degree $\mathbf{l}\oplus 3$ or less such that $\vdash \mathfrak{C}_1 = B^{\mathbf{k}\oplus 2}I \times \mathfrak{C}_1'$ and $\vdash \mathfrak{C}_2 = B^{\mathbf{l}\oplus 2}I \times \mathfrak{C}_2'$. Hence both $B^{\mathbf{k}\oplus 2}I \times \mathfrak{C}_1'$ and $B^{\mathbf{l}\oplus 2}I \times \mathfrak{C}_2'$ have the degree n and correspond to π. However, they have the degrees $\mathbf{k}\oplus 3$ and $\mathbf{l}\oplus 3$ respectively so that $\mathbf{k} = \mathbf{l}$ and $B^{\mathbf{k}\oplus 2}I = B^{\mathbf{l}\oplus 2}I$. Also $\mathfrak{C}_1' = \mathfrak{C}_2'$ since each corresponds to π. Hence $\mathfrak{C}_1 = \mathfrak{C}_2$.

We note that if a regular combinator $\mathfrak{M}$, which contains at least one term of the form $B^{\mathbf{n}}W$, corresponds to the normal sequence μ, then μ can be broken up in several ways into the product of a duplication sequence, ω, and a permutation sequence, π. The reason for this is that, owing to the action of the duplication sequence, there will be some terms which appear several times and different ways of permuting these terms will give the same normal sequence μ. Corresponding to each of the ways of breaking up μ, there is determined a seminormal combinator which corresponds to μ. We wish to prove these equal to each other. Curry has shown how to do this.[6] His proof uses only theorems which we have proved, together with the theorem

$$\vdash B^{\mathbf{n}}W \times B^{\mathbf{n}}C = B^{\mathbf{n}}W$$

which follows immediately from P15 and T28.

M60. If $\mathfrak{W}$, $\mathfrak{C}$, and $\mathfrak{B}$ are respectively a normal duplicator, a normal permutator, and a normal compositor, and $\mathfrak{W} \times \mathfrak{C} \times \mathfrak{B}$ has the degree $\mathbf{n}\oplus 1$, then $\vdash B^{\mathbf{n}}I \times \mathfrak{W} \times \mathfrak{C} \times \mathfrak{B} = \mathfrak{W} \times \mathfrak{C} \times \mathfrak{B}$.

Throughout these Lemmas let $\mathbf{k}$, $\mathbf{l}$, $\mathbf{m}$, and $\mathbf{n}$ stand for numbers.

LEMMA 1. If $\mathbf{n}$ not conv 0, $B^{\mathbf{m}}B^{\mathbf{n}}$ has the degree $\mathbf{m}\oplus\mathbf{n}\oplus 2$ and if Ω has the

[6] H. B. Curry, *An analysis of logical substitution*, Am. Jour. Math., vol. 51 (1929), Lemma 5, pp. 381–383. We shall refer to this lemma as "Lemma 5, Curry 1929."

degree $\mathbf{k}\oplus 1$, then $B^{\mathbf{m}}B^{\mathbf{n}} \times \Omega$ has the degree $\mathbf{m}\oplus\mathbf{n}\oplus 2$ if $\mathbf{m}\oplus 1 \geqq \mathbf{k}$, and the degree $\mathbf{k}\oplus\mathbf{n}\oplus 1$ if $\mathbf{m}\oplus 1 \leqq \mathbf{k}$.

Obvious by *Satz* 1, II C3, p. 801.

Lemma 2. If $\mathbf{n}$ is not conv 0, $(\Delta\mathbf{n})^{\mathbf{m}}$ has the degree $\mathbf{n}\oplus 2$ and if Ω has the degree $\mathbf{k}\oplus 1$, then $(\Delta\mathbf{n})^{\mathbf{m}} \times \Omega$ has the degree $\mathbf{n}\oplus 2$ if $\mathbf{n}\oplus 1 \geqq \mathbf{k}$, and the degree $\mathbf{k}\oplus 1$ if $\mathbf{n}\oplus 1 \leqq \mathbf{k}$.

Obvious from a proof like that of *Satz* 2, II E2, p. 826.

Lemma 3. If $\mathbf{n}$ is not conv 0, $B^{\mathbf{m}}W^{\mathbf{n}}$ has the degree $\mathbf{m}\oplus 2$ and if Ω has the degree $\mathbf{k}\oplus 1$, then $B^{\mathbf{m}}W^{\mathbf{n}} \times \Omega$ has the degree $\mathbf{m}\oplus 2$ if $\mathbf{m}\oplus\mathbf{n}\oplus 1 \geqq \mathbf{k}$, and the degree $\mathbf{m}\oplus\mathbf{l}\oplus 2$ if $\mathbf{m}\oplus\mathbf{n}\oplus 1 \leqq \mathbf{k}$ and $\mathbf{m}\oplus\mathbf{n}\oplus\mathbf{l}\oplus 1 = \mathbf{k}$.

Obvious by *Satz* 4, II C4, p. 805.

Lemma 4. If $\mathbf{n}$ not conv 0, $\vdash B^{\mathbf{m}}B^{\mathbf{n}} = B^{\mathbf{m}\oplus\mathbf{n}\oplus 1}I \times B^{\mathbf{m}}B^{\mathbf{n}}$, and if Ω conv $B^{\mathbf{k}}I \times \Omega$ then:

1. $B^{\mathbf{m}}B^{\mathbf{n}} \times \Omega$ conv $B^{\mathbf{m}\oplus\mathbf{n}\oplus 1}I \times B^{\mathbf{m}}B^{\mathbf{n}} \times \Omega$ if $\mathbf{m}\oplus 1 \geqq \mathbf{k}$.
2. $B^{\mathbf{m}}B^{\mathbf{n}} \times \Omega$ conv $B^{\mathbf{k}\oplus\mathbf{n}}I \times B^{\mathbf{m}}B^{\mathbf{n}} \times \Omega$ if $\mathbf{m}\oplus 1 \leqq \mathbf{k}$.

Obvious by M54 and M53.

Lemma 5. If $\mathbf{n}$ is not conv 0, $\vdash (\Delta\mathbf{n})^{\mathbf{m}} = B^{\mathbf{n}\oplus 1}I \times (\Delta\mathbf{n})^{\mathbf{m}}$, and if Ω conv $B^{\mathbf{k}}I \times \Omega$ then:

1. $(\Delta\mathbf{n})^{\mathbf{m}} \times \Omega$ conv $B^{\mathbf{n}\oplus 1}I \times (\Delta\mathbf{n})^{\mathbf{m}} \times \Omega$ if $\mathbf{n}\oplus 1 \geqq \mathbf{k}$.
2. $(\Delta\mathbf{n})^{\mathbf{m}} \times \Omega$ conv $B^{\mathbf{k}}I \times (\Delta\mathbf{n})^{\mathbf{m}} \times \Omega$ if $\mathbf{n}\oplus 1 \leqq \mathbf{k}$.

Use T24, F31, T31, M55, and M53.

Lemma 6. If $\mathbf{n}$ is not conv 0, $\vdash B^{\mathbf{m}}W^{\mathbf{n}} = B^{\mathbf{m}\oplus 1}I \times B^{\mathbf{m}}W^{\mathbf{n}}$ and if Ω conv $B^{\mathbf{k}}I \times \Omega$ then:

1. $B^{\mathbf{m}}W^{\mathbf{n}} \times \Omega$ conv $B^{\mathbf{m}\oplus 1}I \times B^{\mathbf{m}}W^{\mathbf{n}} \times \Omega$ if $\mathbf{m}\oplus\mathbf{n}\oplus 1 \geqq \mathbf{k}$.
2. $B^{\mathbf{m}}W^{\mathbf{n}} \times \Omega$ conv $B^{\mathbf{m}\oplus\mathbf{l}\oplus 1}I \times B^{\mathbf{m}}W^{\mathbf{n}} \times \Omega$ if $\mathbf{m}\oplus\mathbf{n}\oplus 1 \leqq \mathbf{k}$ and $\mathbf{m}\oplus\mathbf{n}\oplus 1\oplus\mathbf{l} = \mathbf{k}$.

Use T32, T24, M56, and M53.

If we compare Lemma 1 with Lemma 4, Lemma 2 with Lemma 5, and Lemma 3 with Lemma 6, the proof of the theorem becomes obvious.

M61. If Ω is a seminormal combinator:

a. If Ω and $B^{\mathbf{n}}I \times \Omega$ have the same degree, then $\vdash \Omega = B^{\mathbf{n}}I \times \Omega$.

b. If Ω and $\Omega \times B^{\mathbf{m}}I$ have the same degree, then $\vdash \Omega = \Omega \times B^{\mathbf{m}}I$.

Proof obvious using M53 and the Lemmas of M60.

M62. If we have a regular combinator Ω, then there is a normal combinator $\mathfrak{N}$, such that $\vdash \Omega = \mathfrak{N}$.

Proof. First we use T36. Then we use *Satz* 6, II D3, p. 815. Then we use *Satz* 1, II D3, p. 811. Then we use T36. Then we use T37. Then we use Lemma 5, Curry 1929, in order that the right regular permutator shall appear. Then we use M58. Then we use T36. Then we use M57. Finally we use M61 so that the right normal identity shall appear. Then we have converted Ω into a normal combinator $\mathfrak{N}$. Therefore by M52 $\vdash \Omega = \mathfrak{N}$.

M63. Two regular combinators can be proved equal if, and only if, they correspond with the same order to the same normal sequence.

Proof analogous to that of M59.

This is the nearest to *Satz* 5, II D6, p. 820 that can be proved. It is an essential part of the isomorphism with Church's system that two combinators cannot be proved equal if they have different degrees.

Satz 2, II D6, p. 820 is valid. In fact we have been using it already.

All of II E1 can be used if we make two slight changes:

1. We modify *Satz* 3 to read:

If $\mathbf{X}$ and $\mathbf{Y}$ are arbitrary combinations and $\mathbf{Y}$ conv $1\mathbf{Y}$, then $\mathbf{XY}$ conv $(\mathbf{X} \times B\mathbf{Y})I$.

We have to use M40 in the proof.

2. We replace *Satz* 6 by:

M64. For every sequence of proper symbols there is a normal $\mathfrak{R}_1$, of minimum degree $\mathbf{n}\oplus 1$ and with no factor of the form $B^{\mathbf{k}}$, such that $(\mathfrak{R}_1 I)$ corresponds to the sequence with order $\mathbf{n}$. Moreover if $\mathfrak{R}_2$ is a normal combinator and $(\mathfrak{R}_2 I)$ corresponds to the sequence with order $\mathbf{m}$ then either $\vdash \mathfrak{R}_2 = B^{\mathbf{m}}I \times \mathfrak{R}_1$ or else there is a $\mathbf{k}$ such that $\vdash \mathfrak{R}_2 = B^{\mathbf{m}}I \times \mathfrak{R}_1 \times B^{\mathbf{k}}$.

Proof. We follow the proof of *Satz* 6 to show that $\mathfrak{R}_1$ can be found.

Now suppose that $(\mathfrak{R}_2 I)$ corresponds with the order $\mathbf{m}$ to the sequence. If $\mathfrak{R}_2$ does not have any factors of the form $B^{\mathbf{k}}$, then $\mathfrak{R}_1$ and $\mathfrak{R}_2$ correspond to the same sequence and $\vdash \mathfrak{R}_2 = B^{\mathbf{m}}I \times \mathfrak{R}_1$ by M63. Consider that $\vdash \mathfrak{R}_2 = \mathfrak{R}_2' \times B^{\mathbf{k}}$ where $\mathfrak{R}_2'$ does not have any factors of the form $B^{\mathbf{k}}$. Then $\mathfrak{R}_2'$ and $\mathfrak{R}_1$ correspond to the same sequence and if $\mathfrak{R}_2'$ is of degree $\mathbf{l}\oplus 1$, $\vdash \mathfrak{R}_2' = B^{\mathbf{l}}I \times \mathfrak{R}_1$. Hence $\vdash \mathfrak{R}_2 = \mathfrak{R}_2' \times B^{\mathbf{k}}, = (B^{\mathbf{l}}I \times \mathfrak{R}_1) \times B^{\mathbf{k}}, = B^{\mathbf{m}}I \times \mathfrak{R}_1 \times B^{\mathbf{k}}$ by M60.

II E2 is completely valid.

Definition. If $B\mathbf{qrs}$ is part of $\mathbf{p}$, the act of replacing a single occurrence of $B\mathbf{qrs}$ in $\mathbf{p}$ by $\mathbf{q(rs)}$ is called a *B-reduction*. The order of $B\mathbf{qrs}$ in $\mathbf{p}$ is called the *order* of the B-reduction.

Definition. If $C\mathbf{qrs}$ is part of $\mathbf{p}$, the act of replacing a single occurrence of $C\mathbf{qrs}$ in $\mathbf{p}$ by $\mathbf{qsr}$ is called a *C-reduction*. The order of $C\mathbf{qrs}$ in $\mathbf{p}$ is called the *order* of the C-reduction.

Definition. If $W\mathbf{qr}$ is part of $\mathbf{p}$, the act of replacing a single occurrence of $W\mathbf{qr}$ in $\mathbf{p}$ by $\mathbf{qrr}$ is called a *W-reduction*. The order of $W\mathbf{qr}$ in $\mathbf{p}$ is called the *order* of the W-reduction.

The word "reduction" alone shall still refer to I- or J-reduction.

Satz 1 and *Satz* 2, II E3, p. 828 are valid if we make the following changes in the statement of them: 1. Replace K by I wherever it appears. 2. Replace "reduction" by "*B*-, *C*-, *W*-, or *I*-reduction" with corresponding changes in the meaning of "reducible."

We note the following additional changes to be made in the proof of *Satz* 2.

Suppose that $\mathfrak{R}$ is of degree $m + 1$. It is obviously possible to choose Ω (line 3, p. 829) so that its degree is the smaller of the two numbers, $p + 1$ and $m + 1$. We do not use K, therefore we must have $q + 1 \geqq p$. We choose $\mathfrak{R}_1$ of degree $m + q + 2 - p$. Then $\Omega \times \mathfrak{R}_1$ is of degree $m + 1$. Now suppose that $\mathfrak{R}$ is of the form $\mathfrak{R}^* \times B^{\mathbf{k}}$ where no factors of the form $B^{\mathbf{k}}$ occur in $\mathfrak{R}^*$. Then $\vdash \mathfrak{R}I = (\mathfrak{R}^* \times B^{\mathbf{k}})I, = \mathfrak{R}^*(B^{\mathbf{k}}I), = (\mathfrak{R}^* \times B^{\mathbf{k}\oplus 1}I)I$ by T27d and *Satz* 3, II E1, .p 823.

Now by Lemmas 1 to 3 of M60, $\mathfrak{R}^* \times B^{\mathbf{k}}$ and $\mathfrak{R}^* \times B^{\mathbf{k}\oplus 1}I$ both have the same degree. Hence $\mathfrak{R}^* \times B^{\mathbf{k}\oplus 1}I$ has the same degree as $\Omega \times \mathfrak{R}_1$. Also $\mathfrak{R}^* \times B^{\mathbf{k}\oplus 1}I$ and $\Omega \times \mathfrak{R}_1$ correspond to the same sequence. Hence, by M63, $\vdash \mathfrak{R}^* \times B^{\mathbf{k}\oplus 1}I = \Omega \times \mathfrak{R}_1$. Hence $(\mathfrak{R}^* \times B^{\mathbf{k}\oplus 1}I)I\mathbf{Y}_1\mathbf{Y}_2 \cdots \mathbf{Y}_p$ conv $(\Omega \times \mathfrak{R}_1)I\mathbf{Y}_1\mathbf{Y}_2 \cdots \mathbf{Y}_p$. However $\vdash (\mathfrak{R}^* \times B^{\mathbf{k}\oplus 1}I)I = \mathfrak{R}I$. Therefore **S** conv $\mathfrak{R}_1 I\mathbf{Y}_0'\mathbf{Y}_1' \cdots \mathbf{Y}_q'$.

The rest of the proof is just the same except that in definition (a) under (3) we simply define $\mathfrak{R}'$ as the normal form of $(\mathfrak{R}_2 \times \mathfrak{B}I \times B\mathbf{Y}_0')$, and in line 11, p. 831 change "Y_0' W oder K ist." to "Y_0' W ist, und $m \geqq 0$, wenn Y_0' I ist."

The restriction "wenn ein Reductionsprozess $\cdots$ unmöglich ist" which appears in the definition of Curry's *Reductionsprozess zweiter Art* (see p. 790) is not used in the proof of *Satz* 2 and hence *Satz* 2 is valid for the more general type of reduction for which we are stating it.

Obviously any reference which we shall make to *Satz* 1 or *Satz* 2, II E3 will be to these modified forms of them.

M65. If a combination **X** is of degree $\mathbf{n}\oplus 1$, then **X** conv $B^{\mathbf{n}}I \times \mathbf{X}$.

By the definition of degree, there is a **Y** whose leading term is a proper symbol, such that $\mathbf{X}\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_{n+1}$ is reducible in the first sense to **Y**. Now let us replace every J in **X** by $(W \times BC \times B^3C \times B^2B^2)I$. These are p-conversions (see P2), hence if we call the result $\mathbf{X}'$, we have $\mathbf{X}'$ conv **X**. Hence by T16 and Cor. $\mathbf{X}'$ is of degree $\mathbf{n}\oplus 1$, so that $\mathbf{X}'\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_{n+1}$ is reducible in the first sense to $\mathbf{Y}'$ whose leading term is a proper symbol. Now suppose we apply I-, B-, C-, and W-reductions of order one to $\mathbf{X}'\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_{n+1}$. By referring to the definition of B, C, and W, we see that every B-, C-, or W-reduction of order one is equivalent to a series of I- and J-reductions of which at least one is of order one. But by T9, the number of I- and J-reductions of order one is limited and so the number of I-, B-, C-, and W-reductions of order one is limited. Let us apply all possible I-, B-, C-, and W-reductions of order one to $\mathbf{X}'\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_{n+1}$ calling the result $\mathbf{Y}''$.

Lemma. The leading term of $\mathbf{Y}''$ is a proper symbol.

Proof. The leading term is either I, J, or a proper symbol. If the leading term is I, then $\mathbf{Y}''$ must be I, else it would be possible to perform an I-reduction of order one. Hence $\mathbf{Y}''$ is of degree one, contrary to the hypothesis that it is convertible into $\mathbf{X}'\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_{n+1}$ which is of degree zero. Since all the J's which occur in $\mathbf{X}'$ occur as parts of B, C, or W, all J's which occur in $\mathbf{Y}''$ must occur as parts of B, C, or W. Hence if the leading term of $\mathbf{Y}''$ is J, $\mathbf{Y}''$ must have one of the forms B, $B\mathbf{p}$, $B\mathbf{pq}$, C, $C\mathbf{p}$, $C\mathbf{pq}$, W, or $W\mathbf{p}$. However in any of these cases the degree of $\mathbf{Y}''$ would not be zero.

Suppose that after the first I-, B-, C-, or W-reduction of order one on $\mathbf{X}'\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_{n+1}$ the result is $\mathbf{Z}_1$, after the second, the result is $\mathbf{Z}_2$, and so on until we reach $\mathbf{Z}_k$ which will be $\mathbf{Y}''$. Now since $\mathbf{X}'$ is of degree $\mathbf{n}\oplus 1$ there must be a reduction which would have been impossible if we had started with $\mathbf{X}'\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_n$. Suppose that the first such one is the α^{th}. Then $\mathbf{Z}_{\alpha-1}$ must have one of the forms $I\mathbf{x}_{n+1}$, $W\mathbf{p}\mathbf{x}_{n+1}$, $B\mathbf{pq}\mathbf{x}_{n+1}$, or $C\mathbf{pq}\mathbf{x}_{n+1}$ and $\mathbf{Z}_\alpha$ must correspond-

ingly have one of the forms $\mathbf{x}_{n+1}$, $\mathbf{p}\mathbf{x}_{n+1}\mathbf{x}_{n+1}$, $\mathbf{p}(\mathbf{q}\mathbf{x}_{n+1})$, $\mathbf{p}\mathbf{x}_{n+1}\mathbf{q}$. Now (see *Satz* 1, II E3, p. 828) we can find a normal combinator $\mathfrak{R}$ such that

$$\mathfrak{R}I\mathbf{Y}_1\mathbf{Y}_2 \cdots \mathbf{Y}_p\mathbf{x}_{n+2}\mathbf{x}_{n+3} \cdots \mathbf{x}_{n+1+m},$$

where the $\mathbf{Y}_i$'s are B, C, W, or I, reduces formally by B-, C-, W-, or I-reductions to $\mathbf{X}'$. Hence $\mathfrak{R}I\mathbf{Y}_1\mathbf{Y}_2 \cdots \mathbf{Y}_p\mathbf{x}_{n+2}\mathbf{x}_{n+3} \cdots \mathbf{x}_{n+1+m}\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_{n+1}$ reduces formally by B-, C-, W-, or I-reductions to $\mathbf{X}'\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_{r+1}$. Hence the hypothesis of *Satz* 2, II E3 is satisfied. Therefore, denoting $\mathfrak{R}I\mathbf{Y}_1\mathbf{Y}_2 \cdots \mathbf{Y}_p$ by $\mathbf{S}$, we find a series of $\mathbf{S}_i$'s such that:

(1) $\mathbf{S}_i\mathbf{x}_{n+2}\mathbf{x}_{n+3} \cdots \mathbf{x}_{n+1+m}\,\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_{n+1}$ reduces formally by B-, C-, W-, or I-reductions to $\mathbf{Z}_i$.

(2) $\mathbf{S}_i$ is $\mathfrak{R}_iI\mathbf{Y}_1^{(i)}\,\mathbf{Y}_2^{(i)} \cdots \mathbf{Y}_{p_i}^{(i)}$ where the $\mathbf{Y}$'s are B, C, W, or I.

(3) $\mathbf{S}$ conv $\mathbf{S}_1$ conv $\mathbf{S}_2$ conv $\cdots$ conv $\mathbf{S}_i$.

Hence $\mathfrak{R}_\alpha I\mathbf{Y}_1^{(\alpha)}\,\mathbf{Y}_2^{(\alpha)} \cdots \mathbf{Y}_{p_\alpha}^{(\alpha)}\,\mathbf{x}_{n+2}\mathbf{x}_{n+3} \cdots \mathbf{x}_{n+1+m}\,\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_{n+1}$ reduces formally by B-, C-, W-, or I-reductions to $\mathbf{Z}_\alpha$. But from the form of $\mathbf{Z}_\alpha$ it is evident that $\mathfrak{R}_\alpha$ must be of degree $\mathbf{p}_\alpha\oplus 1\oplus\mathbf{m}\oplus\mathbf{n}\oplus 1$. But $\mathfrak{R}_\alpha$ is a normal combinator, so by M61 $\vdash B^{\mathbf{p}_\alpha\oplus 1\oplus\mathbf{m}\oplus\mathbf{n}}I \times \mathfrak{R}_\alpha = \mathfrak{R}_\alpha$.

Hence $\mathbf{S}_\alpha$ conv $\mathfrak{R}_\alpha I\mathbf{Y}_1^{(\alpha)}\,\mathbf{Y}_2^{(\alpha)} \cdots \mathbf{Y}_{p_\alpha}^{(\alpha)}$

conv $(B^{\mathbf{p}_\alpha\oplus 1\oplus\mathbf{m}\oplus\mathbf{n}}I \times \mathfrak{R}_\alpha)I\mathbf{Y}_1^{(\alpha)}\,\mathbf{Y}_2^{(\alpha)} \cdots \mathbf{Y}_{p_\alpha}^{(\alpha)}$

conv $(B^{\mathbf{p}_\alpha} \times B \times B^{\mathbf{m}\oplus\mathbf{n}})I(\mathfrak{R}_\alpha I)\mathbf{Y}_1^{(\alpha)}\,\mathbf{Y}_2^{(\alpha)} \cdots \mathbf{Y}_{p_\alpha}^{(\alpha)}$

conv $B^{\mathbf{m}\oplus\mathbf{n}}I \times (\mathfrak{R}_\alpha I\mathbf{Y}_1^{(\alpha)}\,\mathbf{Y}_2^{(\alpha)} \cdots \mathbf{Y}_{p_\alpha}^{(\alpha)})$

conv $B^{\mathbf{m}\oplus\mathbf{n}}I \times \mathbf{S}_\alpha$. Therefore $\mathbf{S}$ conv $B^{\mathbf{m}\oplus\mathbf{n}}I \times \mathbf{S}$.

Hence $\mathbf{X}$ conv $\mathbf{X}'$

conv $\mathbf{S}\mathbf{x}_{n+2}\mathbf{x}_{n+3} \cdots \mathbf{x}_{n+1+m}$

conv $(B^{\mathbf{m}\oplus\mathbf{n}}I \times \mathbf{S})\mathbf{x}_{n+2}\mathbf{x}_{n+3} \cdots \mathbf{x}_{n+1+m}$

conv $B^{\mathbf{n}}I \times (\mathbf{S}\mathbf{x}_{n+2}\mathbf{x}_{n+3} \cdots \mathbf{x}_{n+1+m})$

conv $B^{\mathbf{n}}I \times \mathbf{X}'$

conv $B^{\mathbf{n}}I \times \mathbf{X}$.

Corollary If $\mathbf{X}$ is of degree $\mathbf{n}$, then $\mathbf{X}$ conv $B^{\mathbf{n}}I\mathbf{X}$.

M66. If $\mathbf{X}$ is a combinator such that $\mathbf{X}\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_n$ reduces by I-, B-, C-, and W-reductions to a combination of proper symbols only, then there is an $\mathfrak{R}$ such that $\mathbf{X}$ conv $\mathfrak{R}I$ and $\mathfrak{R}I\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_n$ reduces by I-, B-, C-, and W-reductions to the same combination of proper symbols.

Proof like that of *Satz* 3, II E3, p. 831. We note that since K's do not appear, $\mathbf{S}_m$ (see line 3, p. 832) cannot contain any $\mathbf{Y}_i$'s since these cannot disappear. Hence $\mathbf{S}_m$ is the same as $\mathfrak{R}_mI$.

M67. If for two combinators $\mathbf{X}$ and $\mathbf{Y}$ there is a k such that $\mathbf{X}\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_k$ and $\mathbf{Y}\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_k$ can be reduced to the same combination of proper symbols by I-, B-, C-, and W-reductions, and if the degrees of $\mathbf{X}$ and $\mathbf{Y}$ are the same, then $\vdash \mathbf{X} = \mathbf{Y}$.

We readily prove $\vdash E\mathbf{X}$ (see proof of M52). Also $\vdash E\mathbf{Y}$. By M66 there are normal combinators $\mathfrak{R}_1$ and $\mathfrak{R}_2$ such that $\vdash \mathbf{X} = \mathfrak{R}_1 I$ and $\vdash \mathbf{Y} = \mathfrak{R}_2 I$. By T16 cor., $\mathfrak{R}_1 I$ and $\mathfrak{R}_2 I$ must have the same degree. If neither $\mathfrak{R}_1$ nor $\mathfrak{R}_2$ contain factors of the form $B^{\mathbf{k}}$, then $\mathfrak{R}_1$ and $\mathfrak{R}_2$ must correspond with the same order to the same sequence and so $\vdash \mathfrak{R}_1 = \mathfrak{R}_2$ by M63. Otherwise, let $\vdash \mathfrak{R}_1 = \mathfrak{R}_1' \times B^{\mathbf{k}}$ and $\vdash \mathfrak{R}_2 = \mathfrak{R}_2' \times B^{\mathbf{l}}$ where neither $\mathfrak{R}_1'$ nor $\mathfrak{R}_2'$ contain factors of the form $B^{\mathbf{k}}$. Then $\vdash \mathfrak{R}_1 I = (\mathfrak{R}_1' \times B^{\mathbf{k}})I = \mathfrak{R}_1'(B^{\mathbf{k}}I) = (\mathfrak{R}_1' \times B^{\mathbf{k}\oplus 1}I)I$, and

$$\vdash \mathfrak{R}_2 I = (\mathfrak{R}_2' \times B^{\mathbf{l}\oplus 1}I)I\,.$$

Hence $\mathfrak{R}_1' \times B^{\mathbf{k}\oplus 1}I$ and $\mathfrak{R}_2' \times B^{\mathbf{l}\oplus 1}I$ correspond with the same order to the same sequence, so by M63, $\vdash \mathfrak{R}_1 I = \mathfrak{R}_2 I$. Then $\vdash \mathbf{X} = \mathbf{Y}$.

M68. If $\mathbf{X}$ and $\mathbf{Y}$ are combinations and $\mathfrak{R}_1$ and $\mathfrak{R}_2$ are regular combinators such that:

1. $\mathbf{X}$ conv $\mathbf{Y}$.
2. Every formula that appears in the conversion from $\mathbf{X}$ to $\mathbf{Y}$, including $\mathbf{X}$ and $\mathbf{Y}$, contains $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$, I, and J.
3. $\mathfrak{R}_1 I\mathbf{y}_1\mathbf{y}_2\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_n$ and $\mathfrak{R}_2 I\mathbf{y}_1\mathbf{y}_2\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_n$ reduce by I-, B-, C-, and W-reductions to $\mathrm{S}^{I\,J}_{\mathbf{y}_1\mathbf{y}_2}\mathbf{X}\,|$ and $\mathrm{S}^{I\,J}_{\mathbf{y}_1\mathbf{y}_2}\mathbf{Y}\,|$ (see Kleene, p. 533).
4. Both $\mathfrak{R}_1$ and $\mathfrak{R}_2$ are of degree $n + 3$; then $\mathfrak{R}_1 IIJ$ conv $\mathfrak{R}_2 IIJ$.

Cf. Theorem 4, Curry 1932.

Throughout the proof we will use "H-reducible" to mean "reducible by I-, B-, C-, and W-reductions." Also we shall let $\mathfrak{X}$ stand for $\mathrm{S}^{I\,J}_{\mathbf{y}_1\mathbf{y}_2}\mathbf{X}\,|$ and $\mathfrak{Y}$ stand for $\mathrm{S}^{I\,J}_{\mathbf{y}_1\mathbf{y}_2}\mathbf{Y}\,|$.

Proof. It is obviously sufficient to prove the theorem for the case that $\mathbf{X}$ goes into $\mathbf{Y}$ by a single conversion.

Case 1. $\mathbf{X}$ goes into $\mathbf{Y}$ by a p-conversion. Then $\mathbf{r} = \mathbf{s}$ or $\mathbf{s} = \mathbf{r}$ is a conversion postulate and $\mathbf{r}$ is part of $\mathbf{X}$ and $\mathbf{Y}$ is the result of replacing a single occurrence of $\mathbf{r}$ in $\mathbf{X}$ by $\mathbf{s}$. By inspection the only undefined terms that appear in $\mathbf{r}$ and $\mathbf{s}$ are $I, J,($, and $)$. Let $\mathbf{r}'$ and $\mathbf{s}'$ stand for $\mathrm{S}^{I\,J}_{\mathbf{y}_1\mathbf{y}_2}\mathbf{r}\,|$ and $\mathrm{S}^{I\,J}_{\mathbf{y}_1\mathbf{y}_2}\mathbf{s}\,|$. Then clearly we can choose a regular combinator $\mathbf{T}$ such that either:

(a) $\mathbf{T}I\mathbf{r}'\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_n$ and $\mathbf{T}I\mathbf{s}'\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_n$ are H-reducible to $\mathfrak{X}$ and $\mathfrak{Y}$ and $\mathbf{T}$ is of degree $n + 2$.

(b) $\mathbf{T}I\mathbf{r}'\mathbf{y}_1\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_n$ and $\mathbf{T}I\mathbf{s}'\mathbf{y}_1\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_n$ are H-reducible to $\mathfrak{X}$ and $\mathfrak{Y}$ and $\mathbf{T}$ is of degree $n + 3$.

(c) $\mathbf{T}I\mathbf{r}'\mathbf{y}_2\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_n$ and $\mathbf{T}I\mathbf{s}'\mathbf{y}_2\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_n$ are H-reducible to $\mathfrak{X}$ and $\mathfrak{Y}$ and $\mathbf{T}$ is of degree $n + 3$.

(d) $\mathbf{T}I\mathbf{r}'\mathbf{y}_1\mathbf{y}_2\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_n$ and $\mathbf{T}I\mathbf{s}'\mathbf{y}_1\mathbf{y}_2\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_n$ are H-reducible to $\mathfrak{X}$ and $\mathfrak{Y}$ and $\mathbf{T}$ is of degree $n + 4$.

We can also find regular combinators $\mathbf{U}$ and $\mathbf{V}$ such that either:

(e) $\mathbf{U}I\mathbf{y}_1$, and $\mathbf{V}I\mathbf{y}_1\mathbf{y}_2$ are H-reducible to $\mathbf{r}'$ and $\mathbf{s}'$.

(f) $\mathbf{U}I\mathbf{y}_1\mathbf{y}_2$ and $\mathbf{V}I\mathbf{y}_1$ are H-reducible to $\mathbf{r}'$ and $\mathbf{s}'$.

(g) $\mathbf{U}I\mathbf{y}_2$ and $\mathbf{V}I\mathbf{y}_1\mathbf{y}_2$ are H-reducible to $\mathbf{r}'$ and $\mathbf{s}'$.

(h) $\mathbf{U}I\mathbf{y}_1\mathbf{y}_2$ and $\mathbf{V}I\mathbf{y}_2$ are H-reducible to $\mathbf{r}'$ and $\mathbf{s}'$.

(i) $\mathbf{U}I\mathbf{y}_1\mathbf{y}_2$ and $\mathbf{V}I\mathbf{y}_1\mathbf{y}_2$ are H-reducible to $\mathbf{r}'$ and $\mathbf{s}'$.

The proofs for different combinations of these cases are similar. Take the proof for a typical combination of cases. Let cases (c) and (f) hold. Let $\mathbf{R}$ be a regular combinator of degree 5 corresponding to $\mathbf{x}_0\mathbf{x}_2(\mathbf{x}_1\mathbf{x}_2\mathbf{x}_3\mathbf{x}_4)\mathbf{x}_4\mathbf{x}_5 \cdots$ and $\mathbf{S}$ be a regular combinator of degree 5 corresponding to $\mathbf{x}_0\mathbf{x}_2(\mathbf{x}_1\mathbf{x}_2\mathbf{x}_3)\mathbf{x}_4\mathbf{x}_5 \cdots$.

Then $\mathbf{RTU}I$ and $\mathbf{STV}I$ both have degree $n + 2$ as do also $\mathfrak{R}_1I$ and $\mathfrak{R}_2I$. Also $\mathbf{RTU}I\mathbf{y}_1\mathbf{y}_2\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_n$ and $\mathbf{STV}I\mathbf{y}_1\mathbf{y}_2\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_n$ are H-reducible to $\mathfrak{X}$ and $\mathfrak{Y}$. Hence, by M67 $\vdash \mathfrak{R}_1I = \mathbf{RTU}I$ and $\vdash \mathfrak{R}_2I = \mathbf{STV}I$. Hence $\mathfrak{R}_1IIJ$ conv $\mathbf{RTU}IIJ$, conv $\mathbf{T}I\mathbf{r}J$, conv $\mathbf{T}I\mathbf{s}J$, conv $\mathbf{STV}IIJ$, conv $\mathfrak{R}_2IIJ$.

Case 2. $\mathbf{X}$ goes into $\mathbf{Y}$ by a single reduction. Let $\mathbf{U}$ be the combinator concerned in this reduction. Then $\mathbf{U}$ is I or J. Let $\mathfrak{X}'$ be the expression obtained by replacing $\mathbf{U}$ by $\mathbf{y}$, all other I's by $\mathbf{y}_1$, and all other J's by $\mathbf{y}_2$ in $\mathbf{X}$. Let $\mathbf{T}$ be a regular combinator of degree $n + 4$ such that $\mathbf{T}I\mathbf{y}\mathbf{y}_1\mathbf{y}_2\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_n$ is H-reducible to $\mathfrak{X}'$.

If $\mathbf{U}$ is I, let $\mathbf{V}$ be a regular combinator of degree 3 that corresponds to $\mathbf{x}_0\mathbf{x}_1\mathbf{x}_1\mathbf{x}_2\mathbf{x}_3 \cdots$. Then $(\mathbf{V} \times \mathbf{T})I$ is of degree $n + 2$ and $(\mathbf{V} \times \mathbf{T})I\mathbf{y}_1\mathbf{y}_2\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_n$ is H-reducible to $\mathfrak{X}$. Hence $\vdash \mathfrak{R}_1I = (\mathbf{V} \times \mathbf{T})I$. However $\mathbf{T}I\mathbf{U}\mathbf{y}_1\mathbf{y}_2\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_n$ is H-reducible to $\mathfrak{Y}$ and $\mathbf{T}I\mathbf{U}$ is of degree $n + 2$. Hence $\vdash \mathfrak{R}_2I = \mathbf{T}I\mathbf{U}$. Also $\mathbf{V}(\mathbf{T}I)IJ = \mathbf{T}I\mathbf{U}IJ$.

If $\mathbf{U}$ is J, let $\mathbf{V}$ be a regular combinator of degree 3 that corresponds to $\mathbf{x}_0\mathbf{x}_2\mathbf{x}_1\mathbf{x}_2\mathbf{x}_3 \cdots$. Then $(\mathbf{V} \times \mathbf{T})I$ is of degree $n + 2$ and $(\mathbf{V} \times \mathbf{T})I\mathbf{y}_1\mathbf{y}_2\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_n$ is H-reducible to $\mathfrak{X}$. Hence $\vdash \mathfrak{R}_1I = (\mathbf{V} \times \mathbf{T})I$. However $\mathbf{T}I((W \times BC \times B^3C \times B^2B^2)I)\mathbf{y}_1\mathbf{y}_2\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_n$ is H-reducible to $\mathfrak{Y}$ and $\mathbf{T}I((W \times BC \times B^3C \times B^2B^2)I)$ is of degree $n + 2$. Hence $\vdash \mathfrak{R}_2I = \mathbf{T}I((W \times BC \times B^3C \times B^2B^2)I)$, $= \mathbf{T}I\mathbf{U}$ (by P2). Also $\mathbf{V}(\mathbf{T}I)IJ = \mathbf{T}I\mathbf{U}IJ$.

Therefore $\mathfrak{R}_1IIJ$ conv $(\mathbf{V} \times \mathbf{T})IIJ$, conv $\mathbf{V}(\mathbf{T}I)IJ$, conv $\mathbf{T}I\mathbf{U}IJ$, conv $\mathfrak{R}IIJ$.

Case 3. $\mathbf{X}$ goes into $\mathbf{Y}$ by a single expansion. Then $\mathbf{Y}$ goes into $\mathbf{X}$ by a single reduction and we use case 2.

MV. If $\mathbf{X}$ and $\mathbf{Y}$ are combinations of the same degree, m, and neither $\mathbf{X}$ nor $\mathbf{Y}$ contain $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_m$, and $\mathbf{X}\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_m$ conv $\mathbf{Y}\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_m$, then $\mathbf{X}$ conv $\mathbf{Y}$.

Proof.

$$\begin{aligned} B^{\mathrm{m}}(I \times I)\mathbf{X}\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_m &\text{ conv } (I \times I)(\mathbf{X}\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_m) \\ &\text{ conv } (I \times I)(\mathbf{Y}\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_m) \\ &\text{ conv } B^{\mathrm{m}}(I \times I)\mathbf{Y}\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_m. \end{aligned}$$

Now suppose that $\mathbf{x}_{m+1}, \mathbf{x}_{m+2}, \cdots, \mathbf{x}_{m+n}$ are the proper symbols occurring in $\mathbf{X}$. Then they must be the same as the proper symbols that appear in $\mathbf{Y}$ since K does not appear in our system. Then we have $B^{\mathbf{m}}(I \times I)\mathbf{X}\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_m$ conv $B^{\mathbf{m}}(I \times I)\mathbf{Y}\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_m$ and every formula that appears in the conversion contains $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_{m+n}$, I, and J.

Now choose $\mathfrak{R}_1$ and $\mathfrak{R}_2$ regular combinators so that $\mathfrak{R}_1 I\mathbf{y}_1\mathbf{y}_2\mathbf{x}_{m+1}\mathbf{x}_{m+2} \cdots \mathbf{x}_{m+n}$ and $\mathfrak{R}_2 I\mathbf{y}_1\mathbf{y}_2\mathbf{x}_{m+1}\mathbf{x}_{m+2} \cdots \mathbf{x}_{m+n}$ are H-reducible to $\overset{I\ J}{\underset{\mathbf{y}_1\mathbf{y}_2}{\mathrm{S}}}\ B^{\mathbf{m}}(I \times I)\mathbf{X}\Big|$ and $\overset{I\ J}{\underset{\mathbf{y}_1\mathbf{y}_2}{\mathrm{S}}}\ B^{\mathbf{m}}(I \times I)\mathbf{Y}\Big|$. Then $(B^{\mathbf{m}\oplus\mathbf{n}\oplus 2}I \times \mathfrak{R}_1)I\mathbf{y}_1\mathbf{y}_2\mathbf{x}_{m+1}\mathbf{x}_{m+2} \cdots \mathbf{x}_{m+n}\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_m$ and $(B^{\mathbf{m}\oplus\mathbf{n}\oplus 2}I \times \mathfrak{R}_2)I\mathbf{y}_1\mathbf{y}_2\mathbf{x}_{m+1}\mathbf{x}_{m+2} \cdots \mathbf{x}_{m+n}\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_m$ are H-reducible to $\overset{I\ J}{\underset{\mathbf{y}_1\mathbf{y}_2}{\mathrm{S}}}\ B^{\mathbf{m}}(I \times I)\mathbf{X}\Big|\,\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_m$ and $\overset{I\ J}{\underset{\mathbf{y}_1\mathbf{y}_2}{\mathrm{S}}}\ B^{\mathbf{m}}(I \times I)\mathbf{Y}\Big|\,\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_m$. Moreover $B^{\mathbf{m}\oplus\mathbf{n}\oplus 2}I \times \mathfrak{R}_1$ and $B^{\mathbf{m}\oplus\mathbf{n}\oplus 2}I \times \mathfrak{R}_2$ are both of degree $\mathbf{m}\oplus\mathbf{n}\oplus 3$. Therefore, by M68, $(B^{\mathbf{m}\oplus\mathbf{n}\oplus 2}I \times \mathfrak{R}_1)IIJ$ conv $(B^{\mathbf{m}\oplus\mathbf{n}\oplus 2}I \times \mathfrak{R}_2)IIJ$. However $(B^{\mathbf{m}\oplus\mathbf{n}\oplus 2}I \times \mathfrak{R}_1)IIJ\mathbf{x}_{m+1}\mathbf{x}_{m+2} \cdots \mathbf{x}_{m+n}$ is convertible into $B^{\mathbf{m}}I(\mathfrak{R}_1 IIJ\mathbf{x}_{m+1}\mathbf{x}_{m+2} \cdots \mathbf{x}_{m+n})$ which is H-reducible to $B^{\mathbf{m}}I(B^{\mathbf{m}}(I \times I)\mathbf{X})$. Also by M65, Cor., and P5, $\mathbf{X}$ conv $B^{\mathbf{m}}I(B^{\mathbf{m}}(I \times I)\mathbf{X})$. Similarly $(B^{\mathbf{m}\oplus\mathbf{n}\oplus 2}I \times \mathfrak{R}_2)IIJ\mathbf{x}_{m+1}\mathbf{x}_{m+2} \cdots \mathbf{x}_{m+n}$ conv $\mathbf{Y}$. Hence $\mathbf{X}$ conv $\mathbf{Y}$.

MC and MV together provide a very powerful method of proving equality.

M69. If neither $\mathbf{X}$ nor $\mathbf{Y}$ contains $\mathbf{x}_1\mathbf{x}_2, \cdots, \mathbf{x}_m$ and $\mathbf{X}\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_m$ conv $\mathbf{Y}\mathbf{x}_1\mathbf{x}_2 \cdots \mathbf{x}_m$, then $B^{\mathbf{m}}I\mathbf{X}$ conv $B^{\mathbf{m}}I\mathbf{Y}$.

Proof similar to that of MV.

Section G

Definition of the statement "M denotes $\mathbf{M}$".

1. Any combination denotes itself if it contains no I's or J's.
2. If F denotes $\mathbf{F}$ and A denotes $\mathbf{A}$, then $\{F\}(A)$ denotes $(\mathbf{FA})$.
3. If M denotes $\mathbf{M}$, and $\mathbf{M}$ conv $\mathbf{Xx}$, and $\mathbf{x}$ is not a term of $\mathbf{X}$, then $\lambda\mathbf{x}[M]$ shall denote $1\mathbf{X}$.

In order to have a sufficiency of proper symbols we introduce as additional undefined terms all the lower case Roman italics both with and without numerical subscripts. We allow the same abbreviations of these formulas that Church and Kleene do (see Church 1932, pp. 353–354 and Kleene, p. 534). We say that a formula is weakly-well-formed if it is well-formed in the sense of Kleene (p. 530) and does not contain I or J.

M70. If M denotes $\mathbf{M}$, then $\mathbf{M}$ is weakly-well-formed and the free symbols of M are proper symbols of $\mathbf{M}$ and the proper symbols of $\mathbf{M}$ are free symbols of M.

We prove this by induction on the number of proper symbols in M.

1. There must be at least one proper symbol in M. If there is only one, then no other undefined terms occur in M and M and $\mathbf{M}$ are the same and M is a weakly well-formed.

Assume our proposition for the case that there are n or less proper symbols in M. Let there be $n + 1$ proper symbols in M.

2. Let M have the form $\{ F \}(A)$. Then there are $\mathbf{F}$ and $\mathbf{A}$ such that F denotes $\mathbf{F}$, A denotes $\mathbf{A}$, and $\{ F \}(A)$ denotes $(\mathbf{FA})$. The proposition follows readily.

3. Let M have the form $\lambda\mathbf{x}[P]$. Then there are $\mathbf{P}$ and $\mathbf{X}$ such that P denotes $\mathbf{P}$, P conv $\mathbf{Xx}$, $\mathbf{x}$ is not a term of $\mathbf{X}$, and M denotes $\mathbf{1X}$. Then P is weakly-well-formed and all the proper symbols of $\mathbf{P}$ are free symbols of P, and all the free symbols of P are proper symbols of $\mathbf{P}$. But the proper symbols of $\mathbf{P}$ are the same as the proper symbols of $\mathbf{Xx}$, since P conv $\mathbf{Xx}$. Hence, in particular, $\mathbf{x}$ is a free symbol of P, so that $\lambda\mathbf{x}[P]$ is weakly-well-formed. Also the free symbols of $\lambda\mathbf{x}[P]$ are all the free symbols of P except $\mathbf{x}$ and this is the same as all the proper symbols of $\mathbf{1X}$ since $\mathbf{X}$ does not contain $\mathbf{x}$ and $\mathbf{1}$ contains no proper symbols.

M71. If M is weakly-well-formed, it denotes at least one combination.

Proof by induction like that of M70.

The case of one proper symbol, or the case of $n + 1$ proper symbols and M of the form $\{ F \}(A)$ are readily handled.

Let M have the form $\lambda\mathbf{x}[P]$. Then P denotes some combination $\mathbf{P}$ by the hypothesis of the induction. $\mathbf{x}$ must be a proper symbol of $\mathbf{P}$ by M70. Let $\mathbf{x}_1, \mathbf{x}_2, \cdots \mathbf{x}_m$ be the other proper symbols of $\mathbf{P}$ and let $\mathfrak{R}$ be a normal combinator of degree $\leqq m + 4$ such that $\mathfrak{R}I$ corresponds to $\mathrm{S}^{I\ J\ \mathbf{x}}_{\mathbf{x}_{m+1}\mathbf{x}_{m+2}\mathbf{x}_{m+3}} JII\mathbf{P}I \Big|$ $\mathbf{x}_{m+4}\mathbf{x}_{m+5} \cdots$. Then $(\mathfrak{R}I\mathbf{x}_1 \cdots \mathbf{x}_mIJ)\mathbf{x}$ conv $\mathbf{P}$ and $\mathbf{x}$ is not a term of $(\mathfrak{R}I\mathbf{x}_1 \cdots \mathbf{x}_mIJ)$, so that $\lambda\mathbf{x}[P]$ denotes $\mathbf{1}(\mathfrak{R}I\mathbf{x}_1 \cdots \mathbf{x}_mIJ)$.

M72. If M denotes both $\mathbf{N}$ and $\mathbf{P}$, then $\mathbf{N}$ conv $\mathbf{P}$.

Proof by induction like that of M70.

The case of one proper symbol, or the case of $n + 1$ proper symbols and M of the form $\{ F \}(A)$ are readily handled (see Kleene, 2VII, p. 532).

Let M have the form $\lambda\mathbf{x}[Q]$. Then there are $\mathbf{Q}$ and $\mathbf{R}$ and $\mathbf{X}$ and $\mathbf{Y}$ such that Q denotes both $\mathbf{Q}$ and $\mathbf{R}$ and $\mathbf{Q}$ conv $\mathbf{Xx}$ and $\mathbf{R}$ conv $\mathbf{Yx}$, and $\mathbf{x}$ is not a term of $\mathbf{X}$ or $\mathbf{Y}$, and $\mathbf{N}$ is $\mathbf{1X}$ and $\mathbf{P}$ is $\mathbf{1Y}$. Then $\mathbf{Q}$ conv $\mathbf{R}$ by the hypothesis of the induction. Hence $\mathbf{Xx}$ conv $\mathbf{Yx}$. Then $\mathbf{1X}$ conv $\mathbf{1Y}$ by M69.

Because of M72 we find it convenient to say that if M and N are weakly-well-formed, then M conv N if and only if each combination that M denotes is convertible into each combination that N denotes. This is a different kind of conversion from that defined by Church (Church 1932, p. 357) but the two will be shown to be equivalent.

M73. If M denotes $\mathbf{M}$ and $\mathbf{x}$ is not a bound symbol of M and N denotes $\mathbf{N}$ and the free symbols of N are not bound in M, then $\mathrm{S}^{\mathbf{x}}_{N} M \Big|$ denotes $\mathrm{S}^{\mathbf{x}}_{\mathbf{N}} \mathbf{M} \Big|$.

Proof by induction like that of M70.

The case of one proper symbol, or the case of $n + 1$ proper symbols and M of the form $\{ F \}(A)$ are readily handled.

Let M have the form $\lambda\mathbf{y}[P]$. Then there is a **P** and an **X** such that P denotes **P**, and **P** conv **Xy**, and **X** does not contain **y**, and **M** is **1X**.

Then $\mathrm{S}\,{}^{\mathbf{x}}_{N}\,P\Big|$ denotes $\mathrm{S}\,{}^{\mathbf{x}}_{\mathbf{N}}\,\mathbf{P}\Big|$ by the hypothesis of the induction. But we see readily that $\mathrm{S}\,{}^{\mathbf{x}}_{\mathbf{N}}\,\mathbf{P}\Big|$ conv $\mathrm{S}\,{}^{\mathbf{x}}_{\mathbf{N}}\,\mathbf{Xy}\Big|$, so that $\mathrm{S}\,{}^{\mathbf{x}}_{N}\,M\Big|$ denotes $1\left(\mathrm{S}\,{}^{\mathbf{x}}_{\mathbf{N}}\,\mathbf{X}\Big|\right)$, which is the same as $\mathrm{S}\,{}^{\mathbf{x}}_{\mathbf{N}}\,\mathbf{M}\Big|$.

M74. If M and N are weakly-well-formed and M conv N in the sense of Church, then M conv N.

Proof by induction on the number of proper symbols of M.

1. Let M have a single proper symbol. Then either M is the same as N or M conv N by a single application of III. This latter case can be handled by reversing the proof of Case 2b.

Assume our proposition for the case that there are n or less proper symbols in M. Let there be $n + 1$ proper symbols in M.

2. Let M have the form $\{ F \}(A)$. Any conversion on a part of M can be handled by the hypothesis of the induction because of Kleene 2X. Hence we need consider only three cases.

a. M conv N by an application of *III* to the whole of M. This is handled by reversing the proof of Case 2b.

b. M conv N by an application of II to the whole of M. This means that F has the form $\lambda\mathbf{x}\cdot P$, that $\mathbf{x}$ is not bound in P that no free symbols of A are bound in P, and that N is $\mathrm{S}\,{}^{\mathbf{x}}_{A}\,P\Big|$. Then there is a **P**, an **X**, and an **A** such that P denotes **P** and **P** conv **Xx** and **X** does not contain **x** and $\lambda\mathbf{x}\cdot P$ denotes **1X** and A denotes **A**. Then M denotes **1XA**, which is conv **XA**, which is conv $\mathrm{S}^{\mathbf{x}}_{\mathbf{A}}\,\mathbf{P}$, since **X** does not contain **x**. But N denotes $\mathrm{S}^{\mathbf{x}}_{\mathbf{A}}\,\mathbf{P}\Big|$ by M73.

c. M conv N by an application of **I** to the whole of M. This can be considered as an application of **I** on F followed by an application of **I** on A, and hence can be treated by the hypothesis of the induction.

3. Let M have the form $\lambda\mathbf{x}\cdot P$.

a. Let the conversion be on part of M. Then, by Kleene 2XII and Kleene's correction to Rule III (p. 530), it follows that the conversion must affect P only, so that N has the form $\lambda\mathbf{x}\cdot Q$ and P conv Q in Church's sense. Then there is a **P** and a **Q** and an **X** and a **Y** such that P denotes **P** and **P** conv **Xx** and **X** does not contain **x** and $\lambda\mathbf{x}\cdot P$ denotes **1X** and Q denotes **Q** and **Q** conv **Yx** and **Y** does not contain **x** and $\lambda\mathbf{x}\cdot Q$ denotes **1Y**. But **P** conv **Q** by the hypothesis of the induction and so **Xx** conv **Yx** and so **1X** conv **1Y** by M69.

b. Let the conversion be an application of III to the whole of M. This is handled by reversing the proof of Case 2b.

c. Let the conversion be an application of I to the whole of M. This can be considered as a conversion on part of M except in the case where there are no occurrences of $\mathbf{y}$ in M and N is $\lambda\mathbf{y}\cdot \mathrm{S}^{\mathbf{x}}_{\mathbf{y}} P \,|$. Then there is a $\mathbf{P}$ and an $\mathbf{X}$ such that P denotes $\mathbf{P}$ and $\mathbf{P}$ conv $\mathbf{Xx}$ and $\mathbf{X}$ does not contain $\mathbf{x}$ and $\lambda\mathbf{x}\cdot P$ denotes $1\mathbf{X}$. Now let Q be the result of changing all the bound $\mathbf{x}$'s of P to $\mathbf{y}$'s and let Q denote $\mathbf{Q}$. Then P conv Q in the sense of Church (we avoid difficulties with the variables by performing the conversions on parts of P) and so $\mathbf{P}$ conv $\mathbf{Q}$ by the hypothesis of the induction. Now $\mathrm{S}^{\mathbf{x}}_{\mathbf{y}} P \,|$ is the same as $\mathrm{S}^{\mathbf{x}}_{\mathbf{y}} Q \,|$ and $\mathrm{S}^{\mathbf{x}}_{\mathbf{y}} Q \,|$ denotes $\mathrm{S}^{\mathbf{x}}_{\mathbf{y}} \mathbf{Q} \,|$ by M73, so that $\mathrm{S}^{\mathbf{x}}_{\mathbf{y}} P \,|$ denotes $\mathrm{S}^{\mathbf{x}}_{\mathbf{y}} \mathbf{Q} \,|$. Now $\mathbf{X}$ and $\mathbf{P}$ do not contain $\mathbf{y}$ by M70 and so $\mathbf{Q}$ does not contain $\mathbf{y}$ because $\mathbf{P}$ conv $\mathbf{Q}$. Hence $\mathbf{Xy}$ conv $\mathrm{S}^{\mathbf{x}}_{\mathbf{y}} \mathbf{P} \,|$ conv $\mathrm{S}^{\mathbf{x}}_{\mathbf{y}} \mathbf{Q} \,|$ and so $\lambda\mathbf{y}\cdot \mathrm{S}^{\mathbf{x}}_{\mathbf{y}} P \,|$ (i.e., N) denotes $1\mathbf{X}$. But $\lambda\mathbf{x}\cdot P$ also denotes $1\mathbf{X}$.

It is not true that for every combination there is a weakly-well-formed formula which denotes it. Hence we define the λ-representative of a combination as follows:

First rewrite $(\mathbf{FA})$ as $\{\mathbf{F}\}(\mathbf{A})$ throughout the combination. Then replace every occurrence of I by $\lambda x \cdot x$ and every occurrence of J by $\lambda xyzt\cdot x(y, x(t, z))$. The result shall be called the λ-representative of the combination.

M75. Any combination is convertible into its λ-representative.

That is, it is convertible into any combination which its λ-representative denotes.

Proof. $\lambda x \cdot x$ denotes $1I$ and $\lambda xyzt\cdot x(y, x(t, z))$ denotes $1J$. Hence the λ-representative of a combination $\mathbf{M}$ denotes the combination obtained from $\mathbf{M}$ by replacing I by $1I$ and J by $1J$ and this conv $\mathbf{M}$ because $1I$ conv I and $1J$ conv J.

M76. If M denotes $\mathbf{M}$, then M is convertible in the sense of Church into the λ-representative of $\mathbf{M}$.

Proof by induction like that of M70.

The case of one proper symbol, or the case of $n + 1$ proper symbols and M of the form $\{F\}(A)$ are readily handled.

Let M have the form $\lambda\mathbf{x}\cdot P$. Then there is a $\mathbf{P}$ and an $\mathbf{X}$ such that P denotes $\mathbf{P}$, $\mathbf{P}$ conv $\mathbf{Xx}$, $\mathbf{X}$ does not contain $\mathbf{x}$ and M denotes $1\mathbf{X}$. Let P' and X' be the λ-representatives of $\mathbf{P}$ and $\mathbf{X}$ respectively. Now, corresponding to each step in the conversion from $\mathbf{P}$ to $\mathbf{Xx}$, there is a sequence of applications of Church's rules, so that $X'(\mathbf{x})$ conv P' in the sense of Church. But P' conv P in the sense of Church by the hypothesis of the induction. Hence $\lambda\mathbf{x}\cdot X'(\mathbf{x})$ conv $\lambda\mathbf{x}\cdot P$ in the sense of Church. However the λ-representative of $1\mathbf{X}$ conv $\lambda\mathbf{x}\cdot X'(\mathbf{x})$ in the sense of Church.

COROLLARY. If M and N are weakly-well-formed, and M conv N, then M conv N in the sense of Church.

Because we prove readily that if **M** conv **N**, then the λ-representative of **M** conv the λ-representative of **N** in the sense of Church by just paralleling the conversion from **M** to **N**.

This corollary together with M74 completes the proof of the equivalence of the two kinds of conversion. It will be noticed further that because of (2) of the definition of denoting, R1 and R2 are the same as Church's Rules V and IV. Also that P1 conv '$x \cdot x = x$. Hence Church's rules of procedure and the postulate '$x \cdot x = x$ hold as a result of our rules of procedure and sixteen postulates.

Conversely, if we take a system containing Church's rules and the postulate '$x \cdot x = x$, then our rules and sixteen postulates hold if we consider a combination as an abbreviation for the formula which we have called its λ-representative.

The correspondence between the two kinds of conversion can be used in other ways also, as in Kleene §6, where 6V is seen to follow from M71 and M76, and 6VI from M74, M75 and MC.

Section H

We will indicate a set of rules of procedure equivalent to Church's first three rules. The first six are R3 to R8 inclusive. The seventh and eighth are:

If **f**(I**q**), then **fq**.

If **fq**, then **f**(I**q**).

The remaining thirty occur in fifteen pairs, each pair being derived from a conversion postulate as follows: If the postulate is **r** = **s**, then the corresponding rules are to be

If **fr**, then **fs**.

If **fs**, then **fr**.

By paralleling the theorems in Section B we readily prove that **M** conv **N** if and only if **N** can be derived from **M** by the thirty-eight rules just given. Then, by the equivalence proved in Section G, these thirty-eight rules are equivalent to Church's first three.

PRINCETON UNIVERSITY.

LINEAR DIFFERENTIAL EQUATIONS WITH ALMOST PERIODIC COEFFICIENTS

By Robert H. Cameron

Introduction. In this paper we shall deal with the homogeneous system of differential equations

$$\frac{d}{dt}\xi_\mu(t) = \sum_{\nu=1}^{N} \alpha_{\mu,\nu}(t)\xi_\nu(t) \qquad (\mu = 1, \cdots, N)$$

and the non-homogeneous system

$$\frac{d}{dt}\xi_\mu(t) = \beta_\mu(t) + \sum_{\nu=1}^{N} \alpha_{\mu,\nu}(t)\xi_\nu(t) \qquad (\mu = 1, \cdots, N),$$

where the $\alpha_{\mu,\nu}(t)$, $\beta_\mu(t)$ and $\xi_\mu(t)$ are complex a.p. (almost periodic)[1] functions of the real variable t. It is the purpose of this paper to point out the manner in which the a.p. solutions of the above equation depend on the modules[2] of the $\alpha_{\mu,\nu}(t)$ and $\beta_\mu(t)$. We shall be interested in determining the form of those solutions which are a.p., and not in determining conditions under which a.p. solutions exist. Such conditions have already been given in papers by Favard,[3] Bochner,[4] and Cameron.[5]

For the sake of simplicity in notation, we rewrite the above equations in the form

$$D[x(t)] = A(t)\cdot x(t) \tag{1}$$

and

$$D[x(t)] = A(t)\cdot x(t) + b(t), \tag{2}$$

where $x(t)$ and $b(t)$ are N-dimensional vectors having the components $\xi_1(t), \cdots, \xi_N(t)$ and $\beta_1(t), \cdots, \beta_N(t)$ respectively, $A(t)$ is the matrix of the $\alpha_{\mu,\nu}(t)$, and $A\cdot x$

Received February 4, 1935; presented to the American Mathematical Society December 27, 1934. This paper was written while the author was a National Research Fellow.

[1] H. Bohr, *Zur Theorie der fastperiodischen Funktionen*, Acta Mathematica, vol. 45 (1925), pp. 29–127, esp. p. 30.

[2] A module is a set of numbers which is closed under addition and subtraction. The module of an a.p. function is the smallest module which contains all the Fourier exponents of the function.

[3] *Sur les équations différentielles linéaires à coefficients presque-périodiques*, Acta Mathematica, vol. 51 (1928), pp. 31–81.

[4] *Homogeneous systems of differential equations with almost periodic coefficients*, Journal of the London Mathematical Society, vol. 8, pp. 283–288.

[5] *Linear differential equations with almost periodic coefficients*, to appear in the Annals of Mathematics.

denotes the matrix product obtained by regarding x as a matrix of N rows and one column. The vector $b(t)$ and the matrix $A(t)$ are a.p. in the sense of Bochner.[6]

The homogeneous system. The theorem which we are going to prove concerning the homogeneous system (1) is suggested by the facts which we know hold when $A(t)$ is actually periodic with a period P. In this case the linear manifold consisting of all the bounded solutions has (if it is not a zero manifold) a basis, each element of which is of the form $e^{i\lambda t}z(t)$, where $z(t)$ is a periodic vector function with the period P. The general theorem which holds whenever $A(t)$ is a.p. is the following

THEOREM I. *Let $A(t)$ be an a.p. N-rowed square matrix function having the module M. Then the linear manifold $\mathfrak{A}$ of a.p. solutions of* (1) *has (if it is not a zero manifold) a basis, each element of which is of the form*

$$e^{i\lambda t}z(t) ,$$

where λ is real and $z(t)$ is an a.p. vector function whose module is contained in M.

In particular, if all of the solutions of (1) *are a.p., the general solution of* (1) *is*

$$x(t) = c_1e^{i\lambda_1 t}z_1(t) + \cdots + c_Ne^{i\lambda_N t}z_N(t) ,$$

and the least common module of $z_1(t), \cdots , z_N(t)$ is exactly M.

To establish this theorem, consider any element $x(t)$ of $\mathfrak{A}$ having the Fourier expansion

$$x(t) \sim \sum_{n \in \sigma} \alpha_n e^{i\Lambda_n t} ,$$

where none of the coefficients $\alpha_1, \alpha_2, \cdots$ is zero, and σ is either the set of all positive integers, or the set of all positive integers which do not exceed a certain positive integer. Corresponding to each element p of σ, let σ_p be the set of integers n such that $\Lambda_n - \Lambda_p \in M$. Clearly two of these sets σ_p and σ_q either are equal or have no element in common, and therefore there exists a sequence $p_1, p_2, \cdots$ of elements of σ such that $\sigma = \sigma_{p_1} + \sigma_{p_2} + \cdots$ while $\sigma_{p_j}\cdot\sigma_{p_k} = 0$ if $j \neq k$.

We shall now show that to each $p \in \sigma$ there corresponds an element $x^{(p)}(t)$ of $\mathfrak{A}$ having the Fourier series

$$x^{(p)}(t) \sim \sum_{n \in \sigma_p} \alpha_n e^{i\Lambda_n t} . \tag{3}$$

In the first place, if $\sigma_p = \sigma$ the solution $x^{(p)}(t)$ exists and equals $x(t)$. On the other hand, suppose that there exists an element q of σ such that q is not an element of σ_p. Then by the generalized Kronecker theorem[7] on diophantine

[6] *Abstrakte fastperiodische Funktionen*, Acta Mathematica, vol. 61 (1933), pp. 150–184, esp. p. 151.

[7] H. Bohr, *Neuerer Beweis eines allgemeinen Kronecker'schen Approximationsatzes*, Det Kgl. Danske Videnskabernes Selskab, Mathematisk-Fysiske Meddelelser, vol. 6 (1924–25), article 8.

approximation, we can choose a sequence of real numbers $h_1, h_2, \cdots$ such that for each element λ of M,

$$\lim_{j\to\infty} \lambda h_j = 0 \qquad (\text{mod } 2\pi)\,, \tag{4}$$

and such that

$$\lim_{j\to\infty} (\Lambda_q - \Lambda_l)h_j = g \qquad (\text{mod } 2\pi)\,,$$

where

$$g \neq 0 \qquad (\text{mod } 2\pi)\,. \tag{5}$$

Moreover, we can assume that $h_1, h_2, \cdots$ is so chosen that

$$\bar{x}(t) \equiv \lim_{j\to\infty} x(t + h_j) \qquad \text{and} \qquad Q = \lim_{j\to\infty} e^{i\Lambda_p h_j}$$

exist uniformly in t; for by Bochner's theorem[8] on the normality of a.p. functions some subsequence of the originally chosen sequence $h_1, h_2, \cdots$ has this property, and we change our notation so that the subsequence is $h_1, h_2, \cdots$. Thus

$$\bar{x}(t) \sim \sum_{n\,\epsilon\,\sigma} Q\alpha'_n\, e^{i\Lambda_n t}\,,$$

where $\alpha'_n = \alpha_n$ if $n \,\epsilon\, \sigma_p$ and $\alpha'_q = e^{ig}\alpha_q$. Moreover, it follows from (4) that

$$\lim_{j\to\infty} A(t + h_j) = A(t)$$

uniformly in t; and since

$$x(t + h_j) - x(h_j) = \int_0^t A(u + h_j)x(u + h_j)\,du\,,$$

we have

$$\bar{x}(t) - \bar{x}(0) = \int_0^t A(u)\bar{x}(u)\,du\,.$$

Thus $\bar{x}(t)$ is a solution of (1), and being a.p., is an element of $\mathfrak{A}$.

Now it follows from (5) that $e^{ig} - 1 \neq 0$, and we can define

$$x^*(t) \equiv \frac{Qe^{ig}x(t) - \bar{x}(t)}{Q(e^{ig} - 1)}\,.$$

Evidently

$$x^*(t) \sim \sum_{n\,\epsilon\,\sigma} \alpha''_n e^{i\Lambda_n t}\,,$$

where $\alpha''_n = \alpha_n$ if $n \,\epsilon\, \sigma_p$; and $\alpha''_q = 0$. Thus the set of Fourier exponents of $x^*(t)$ is a proper subset of the set of Fourier exponents of $x(t)$, for at least one exponent, Λ_q, now occurs in a term with a zero coefficient. Moreover, the process we have just used to eliminate Λ_q can be applied again to eliminate

[8] *Fastperiodische Funktione* I, Mathematische Annalen, vol. 96 (1927), pp. 119–147.

any term of $x^*(t)$ whose subscript does not belong to σ_p. Let us repeat this process as long as possible. Clearly the set of solutions obtained in this way is linearly independent, and must therefore be finite. But the process can always be continued as long as the last solution obtained has a term whose subscript is not contained in σ_p. Thus when the process terminates, the last solution obtained has the subscripts of all of its terms contained in σ_p. But the process leaves the coefficients of terms for which $n \in \sigma_p$ invariant. Therefore what we obtain is exactly $x^{(p)}(t)$, and it follows that $x^{(p)}(t)$ exists and is a solution of (1). Moreover, $p \in \sigma_p$, and hence $x^{(p)}(t)$ has the non-vanishing term $\alpha_p\, e^{i\Lambda_p t}$. Thus $x^{(p)}(t)$ is a non-trivial solution.

Having established the existence of the non-trivial solutions $x^{(1)}(t)$, $x^{(2)}(t)$, $\cdots$ defined by (3), we next observe that the number m of elements of the sequence $p_1, p_2, \cdots$ is finite, and that

$$x(t) = x^{(p_1)}(t) + x^{(p_2)}(t) + \cdots + x^{(p_m)}(t)\,. \tag{6}$$

The finiteness of m follows from the fact that $x^{(p_1)}(t)$, $x^{(p_2)}(t)$, $\cdots$ are linearly independent, since no two of them have a Fourier exponent in common, and (6) is a direct consequence of the definition of the sequence $p_1, p_2, \cdots$. Moreover, each term of the right member of (6) is the product of an exponential factor $e^{i\Lambda_{p_j} t}$ and an a.p. vector function $e^{-i\Lambda_{p_j} t}x^{(p_j)}(t)$ whose module is contained in M. Thus each element of $\mathfrak{A}$ is a finite sum of functions of the desired form $e^{i\lambda t}z(t)$, and $\mathfrak{A}$ must have a basis consisting of such functions.

It remains to show that in the special case in which all of the solutions are a.p., the least common module M' of $z_1(t), \cdots, z_N(t)$ is not only contained in M but also contains M. This follows from the fact that each of the $\alpha_{\mu,\nu}(t)$ can be expressed as a quotient of two determinants, the denominator of which is the Wronskian $W(t)$ of a fundamental set of solutions, and the numerator of which is obtained by replacing one column of the denominator by the derivative of a column of the denominator. For if we take the solutions $e^{i\lambda_\mu t}z_\mu(t)$ as our fundamental set, the exponentials cancel out and the two determinants are polynomials in the components of $\zeta_{\mu,\nu}(t)$ of the $z_\mu(t)$ and their derivatives $\zeta'_{\mu,\nu}(t)$. Thus $\alpha_{\mu,\nu}(t)$ is the quotient of two a.p. functions which have their modules contained in M', the denominator function being

$$\Delta(t) \equiv W(t)\, e^{-i(\lambda_1+\cdots+\lambda_N)t}\,.$$

But

$$|\,W(t)\,| = c\, e^{\int_0^t \mathfrak{R}\{\alpha_{1,1}(u)+\cdots+\alpha_{N,N}(u)\}\,du}$$

is a.p., and hence by a theorem of Bochner's[9] $\int_0^t \mathfrak{R}\{\alpha_{1,1}(u) + \cdots + \alpha_{N,N}(u)\}\,du$ is bounded, so that $|\,W(t)\,|$ is bounded away from zero. Thus $\Delta(t)$ is bounded

[9] *Remark on the integration of almost periodic functions*, Journal of the London Mathematical Society, vol. 8, pp. 250–254.

away from zero, and $\alpha_{\mu,\nu}(t)$ is a.p. with a module contained in M'. It follows that $M = M'$.

The non-homogeneous system. Passing now to a consideration of the non-homogeneous system (2), we again find that our general result is suggested by the known facts in the periodic case. If $A(t)$ and $b(t)$ are periodic with a period P, and (2) has at least one bounded solution, then (2) has a solution which is periodic with the period P. In the general case we have

THEOREM II. *Let the N-rowed square matrix function $A(t)$ and the vector function $b(t)$ be a.p., and have M as their least common module. Then if the system (2) has an a.p. solution it has a solution which is a.p. and has its module contained in M.*

In particular, if all of the solutions are a.p., the general solution is

$$x(t) = x_0(t) + e^{i\lambda_1 t}z_1(t) + \cdots + e^{i\lambda_N t}z_N(t),$$

where the least common module of $z_1(t), \cdots, z_N(t)$ is the module of $A(t)$, and the least common module of $x_0(t), z_1(t), \cdots, z_N(t)$ is M.

To prove this theorem, consider the homogeneous system

$$D[x(t)] = A(t) \cdot x(t) + b(t)\, x_{N+1}(t), \qquad \frac{d}{dt} x_{N+1}(t) \equiv 0, \tag{7}$$

where $x_{N+1}(t)$ is an unknown scalar function. This system has at least one a.p. solution, for (2) has an a.p. solution $x^*(t)$, and (7) must therefore have the solution $x(t) \equiv x^*(t)$, $x_{N+1}(t) \equiv 1$. But (7) is a homogeneous system of order $N + 1$ of the form (1). Therefore by Theorem I its manifold of a.p. solutions has a basis, each element of which is of the form

$$x(t) \equiv e^{i\lambda t}z(t), \qquad x_{N+1}(t) \equiv e^{i\lambda t}z_{N+1}(t),$$

where λ is real and the vector function $z(t)$ and the scalar function $z_{N+1}(t)$ are a.p. with their modules contained in M. Moreover, for at least one element of this basis $z_{N+1}(t) \not\equiv 0$; for we have shown that (7) has an a.p. solution whose last component is not identically zero. Let this element be $e^{i\lambda_1 t}z^{(1)}(t)$, $e^{i\lambda_1 t}z^{(1)}_{N+1}(t)$. Then it follows from (7) that $e^{i\lambda_1 t}z^{(1)}_{N+1}(t) \equiv k$, where k is a non-zero constant, and that $x^{(1)}(t) \equiv (1/k)\, e^{i\lambda_1 t}z^{(1)}(t)$ is a solution of (2). Hence $z^{(1)}_{N+1}(t) \equiv ke^{-i\lambda_1 t}$, and λ_1 is an element of M. Thus our general theorem is proved, for $x^{(1)}(t)$ is an a.p. solution of (2) whose module is contained in M. Moreover, the statements made in Theorem II concerning the special case in which all the solutions are a.p. obviously holds.

PRINCETON UNIVERSITY AND THE INSTITUTE FOR ADVANCED STUDY.

ON THE WARING PROBLEM WITH POLYNOMIAL SUMMANDS

By M. Gweneth Humphreys

1. Introduction. When Waring, in 1770, made his famous conjecture concerning the representation of all positive integers as sums of powers of positive integers, he also suggested the possibility of a similar representation by polynomials. Very little was done toward examining this until recently. L. E. Dickson[1] has proved by algebraic methods that every integer greater than a stated integer is expressible as a sum of nine values of a cubic polynomial in which the square term does not appear. Supplementing this proof by table work he has proved universal theorems[2] for special cubic polynomials. Universal theorems have also been proved for special cubic polynomials by Frances E. Baker[3] and G. C. Webber.[4]

E. Landau[5] has proved by analytic methods similar to those of Hardy and Littlewood a theorem for polynomials of degree k which corresponds to their first theorem for k-th powers.[6] R. D. James[7] has completed the analysis and proved that every sufficiently large integer (greater than a finite integer not determined) is a sum of nine values of a polynomial of the form

$$a(x^3 - x)/6 + b(x^2 - x)/2 + cx\,,$$

where $a > 0$, $(a, b, c) = 1$, and $a \not\equiv 4c \pmod 8$.

In the present paper, with the analysis as developed by Landau and James, it is proved that, after excluding exceptional cases similar to $a \not\equiv 4c \pmod 8$ above, every sufficiently large integer can be represented by $s \geqq (k - 2)2^{k-1} + 5$ values of a polynomial of degree k when $k = 4, 5, 6$, or 7. If the polynomial satisfies a further condition, this theorem is proved for $k \leqq 28$.

2. Introductory analysis. Let

$$\phi(x) = \beta_k x^k + \beta_{k-1}x^{k-1} + \cdots + \beta_1 x + \beta_0$$

Received April 10, 1935.

[1] L. E. Dickson, Transactions of the American Mathematical Society, vol. 36 (1934), pp. 731–748.

[2] L. E. Dickson, Transactions of the American Mathematical Society, vol. 36 (1934), pp. 1–12.

[3] Frances E. Baker, Dissertation, Chicago, 1934.

[4] G. C. Webber, Transactions of the American Mathematical Society, vol. 36 (1934), pp. 493–510.

[5] E. Landau, *Über die neue Winogradoffsche Behandlung des Waringschen Problems*, Mathematische Zeitschrift, vol. 31 (1930), pp. 319–338.

[6] E. Landau, *Vorlesungen über Zahlentheorie*, Leipzig, 1927.

[7] R. D. James, American Journal of Mathematics, vol. 56 (1934), pp. 303–315.

be a polynomial of degree k with integral coefficients such that $\beta_k > 0$. Let $r_s(n)$ denote the number of integral solutions of

$$n = \sum_{\nu=1}^{s} \phi(x_\nu), \qquad x_\nu \geqq 0.$$

Then E. Landau[5] has proved that

$$\left| r_s(n) - \frac{\Gamma^s(k+1/k)}{\beta_k^{s/k}\,\Gamma(s/k)} \mathfrak{S}\, n^{(s-k)/k} \right| < C_1\, n^{(s-k-\delta)/k} \tag{1}$$

for $s \geqq (k-2)2^{k-1} + 5$, where C_1 and δ are positive constants depending only on $s, \beta_k, \cdots, \beta_0$. The function $\mathfrak{S}$ is called the singular series and is defined as follows. Let r be prime to q,

$$\rho = e^{2\pi i r/q}, \qquad S_\rho = \sum_{h=0}^{q-1} \rho^{\phi(h)}, \qquad A(q) = \sum_{\substack{r=0\\(r,q)=1}}^{q-1} q^{-s} S_\rho^s\, e^{-2\pi i r n/q}.$$

Then

$$\mathfrak{S} = \sum_{q=1}^{\infty} A(q).$$

If $\mathfrak{S} \geqq \eta > 0$, where η is independent of n, then from (1), $r_s(n) > 0$ when

$$n > C_2 = \left[\frac{C_1\, \beta_k^{s/k}\, \Gamma(s/k)}{\eta\, \Gamma^s(k+1/k)}\right]^{k/\delta}. \tag{2}$$

Let $\theta = \theta(p)$ be the highest power of p which divides every coefficient of $\phi'(x)$, and $\gamma = \gamma(p)$ be defined by

$$\gamma = \begin{cases} \theta + 2, & p = 2, \\ \theta + 1, & p > 2. \end{cases}$$

Denote $p^{-\theta}\phi'(x)$ by $\phi_0(x)$. Let $M(m) = M(m, n)$ be the number of integral solutions of

$$\sum_{\nu=1}^{s} \phi(x_\nu) \equiv n \pmod{m}, \qquad 0 \leqq x_\nu < m. \tag{3}$$

For $m = p^l$, let $N(p^l) = N(p^l, n)$ be the number of integral solutions of this congruence in which at least one $\phi_0(x_\nu)$ is prime to p. Such solutions will be called primitive. Then the following lemmas hold.[7]

LEMMA 1. *If $l \geqq \gamma$, then*

$$N(p^l) = p^{(l-\gamma)(s-1)} N(p^\gamma).$$

LEMMA 2. $M(m) = m^{s-1} \sum_{q \mid m} A(q)$.

LEMMA 3. *If p_h denotes the h-th prime, then*

$$\sum_{q \mid p_1^l \cdots p_l^l} A(q) = \prod_{p \leqq p_l} \sum_{q \mid p^l} A(q).$$

3. **The form of polynomials of degree** k. A polynomial[8] $F(y)$ of degree k which represents an integer for every integral value of $y \geqq 0$ represents an integer for every integral value of y. Then set

$$F(y) = a_k \begin{bmatrix} y \\ k \end{bmatrix} + a_{k-1} \begin{bmatrix} y \\ k-1 \end{bmatrix} + \cdots + a_1 \begin{bmatrix} y \\ 1 \end{bmatrix} + a_0 , \tag{4}$$

where $a_k, \cdots, a_0$ are rational, and $\begin{bmatrix} y \\ s \end{bmatrix} = y(y+1) \cdots (y+s-1)/s!$.

Since $F(y)$ is an integer for all integral values of y, it is for $y = 0$. Therefore a_0 is an integer. Similarly it is seen by setting $y = -1, \cdots, -k+1$, that $a_1, \cdots, a_k$ are all integers.

Conversely, if $a_k, \cdots, a_0$ are all integers, $F(y)$ is an integer for all integral values of y.

Therefore every polynomial of degree k in y which represents an integer for all integral values of $y \geqq 0$ can be expressed in form (4) where $a_k, \cdots, a_0$ are integers.

For the following discussion, $a_k > 0$, and a_0 may be taken equal to zero. For if $n = \sum_{\nu=1}^{s} F(y_\nu)$ has an integral solution for $n > C_2$, then $n = \sum_{\nu=1}^{s} P(y_\nu)$ has an integral solution for $n > C_2 - sa_0$, where $P(y) = F(y) - a_0$. It is also assumed that $a_k, \cdots, a_1$ are such that for no prime p is $P(y) \equiv 0 \pmod p$ for all y, for if this were so for some p, all sums of values of $P(y)$ would be multiples of p. This last condition will be referred to as hypothesis I. It implies $(a_k, \cdots, a_1) = 1$.

It will be useful to write also

$$P(y) = (b_k y^k + \cdots + b_1 y)/k!,$$

where

$$\begin{gathered} b_{k-r} = \sum_{q=0}^{r} d_{qr}\, a_{k-q}, \\ d_{00} = 1, \quad d_{rr} = k(k-1) \cdots (k-r+1), \quad d_{0r} = \sum_{n_1, \cdots, n_r = 1}^{k-1} n_1 \cdots n_r \\ (n_i \neq n_j; \quad i, j = 1, \cdots, r), \\ d_{qr} = k(k-1) \cdots (k-q+1) \sum_{n_1, \cdots, n_{r-q} = 1}^{k-q-1} n_1 \cdots n_{r-q} \\ (n_i \neq n_j; \quad i, j = 1, \cdots, r-q; q = 1, \cdots, r-1). \end{gathered} \tag{5}$$

In $P(y)$ set $y = vx + t$ to give

$$Q(x) = A_k x^k + \cdots + A_1 x + A_0,$$

[8] David Hilbert, Mathematische Annalen, vol. 36 (1890), pp. 511, 512.

where

$$(6)\qquad \begin{aligned} A_k &= b_k v^k/k!\,, \\ A_{k-r} &= \left\{ \sum_{l=0}^{r-1} \frac{(k-l)\cdots(k-r+1)}{(r-l)!} b_{k-l} t^{r-l} + b_{k-r} \right\} v^{k-r}/k!\,, \end{aligned}$$

$t \geqq 0$ is a parameter which is assigned integral values depending on $a_k, \cdots, a_1$, and $v \geqq 0$ is an integer having the smallest number of factors p such that $A_k, \cdots, A_0$ are integers for every value of t. For $p > k$, $p \nmid v$.

By hypothesis I, there are values of t such that there is at least one value of x for which $p \nmid Q(x)$. Such values of t will be called *admissible*. If $p \nmid v$, all values of t are admissible.

Since $Q(x)$ has integral coefficients, $\theta(p)$ as defined in §2 can be determined for each p.

4. **Congruence** (3) **for** $p > k$. As in §2, $r_s(n) > 0$ for $n > C_2$, if $\mathfrak{S} \geqq \eta > 0$. To prove that for $s \geqq (k-2)2^{k-1} + 5$, $\mathfrak{S} \geqq \eta > 0$, we prove that for this value of s and for all primes p and integers n, $N(p^\gamma) > 0$. This proof is divided into two parts according as $p > k$ or $p \leqq k$.

For primes $p > k$ we prove

LEMMA 4. *If* $3 < k \leqq 28$, $s \geqq (k-2)2^{k-1} + 5$, *and* $p > k$, *then* $N(p^\gamma) \geqq 1$.

For $p > k$, $p \nmid v$. By hypothesis I, at least one coefficient of $Q(x)$ is not divisible by p. Then $\theta = 0$, $\gamma = 1$, and the congruence (3) becomes

$$\sum_{\nu=1}^{s} Q(x_\nu) \equiv n \pmod{p}\,.$$

L. E. Dickson[9] has proved that for $k \leqq 28$, $p > k$, and $\sigma = s - 1$, the congruence

$$\sum_{\nu=1}^{\sigma} Q(x_\nu) \equiv n \pmod{p}$$

has at least one integral solution.

Not all the coefficients of $Q_0(x)$ are divisible by p, and therefore there is at least one value of x for which $p \nmid Q_0(x)$. Suppose $p \nmid Q_0(x_1)$. Then from the above

$$\sum_{\nu=1}^{\sigma} Q(x_\nu) \equiv n - Q(x_1) \pmod{p}$$

has an integral solution. Thus

$$\sum_{\nu=1}^{s} Q(x_\nu) \equiv n \pmod{p}$$

[9] L. E. Dickson, American Journal of Mathematics, July, 1935.

has a primitive solution, that is,

$$N(p^\gamma) = N(p) \geqq 1 \qquad \text{for} \qquad s \geqq (k-2)2^{k-1} + 5\,,$$

for every n and every $p > k$.

5. **Congruence** (3) **for** $p \leqq k$. When $p \leqq k$, $Q_0(x)$ is congruent mod p to a polynomial of degree $p - 1$ obtained from $Q_0(x)$ by replacing x^p by x. If for every admissible value of t, all the coefficients of the polynomial so obtained are divisible by p, there obviously can be no primitive solutions of (3). We shall say that $P(y)$ satisfies hypothesis II if for each prime $p \leqq k$ there is an admissible value of t such that for some value of x, $p \nmid Q_0(x)$. Hypothesis II may be stated also in terms of the coefficients $A_k, \cdots, A_1$: for each prime $p \leqq k$ there is an admissible value of t for which the congruences

$$\begin{aligned} A_1 &\equiv 2A_2 + (p+1)A_{p+1} + \cdots \\ &\equiv \cdots \equiv pA_p + (2p-1)A_{2p-1} + \cdots \\ &\equiv 0 \qquad\qquad\qquad\qquad\qquad\qquad \pmod{p^{\theta+1}} \end{aligned}$$

are not all satisfied.

The theorems of the paper will be proved for polynomials which satisfy hypotheses I and II. For $k = 4$ and 5 conditions on the coefficients a of $P(y)$ which determine whether hypothesis II is satisfied are stated. For polynomials of higher degrees such conditions are not stated.

For the discussion of congruence (3) in this case we prove two lemmas.

LEMMA 5. *If* $p \nmid Q(x_1)$, $p \nmid Q_0(x_2)$, *then* $N(p^\gamma) \geqq 1$ *for* $s \geqq p^\gamma + 1$ *and every integer* n.

Any integer is congruent mod p^γ to a sum of at most p^γ values $Q(x_1)$. If $x_2 \equiv x_1 \pmod p$, the lemma follows. If $x_2 \not\equiv x_1 \pmod p$, $n - Q(x_2)$ is congruent mod p^γ to a sum of at most p^γ values $Q(x_1)$, and the lemma follows.

LEMMA 6. *If* $p \nmid (A_k, \cdots, A_1)$, $p \nmid Q(x_1)$, $p \nmid Q_0(x_2)$, *then* $N(p^\gamma) \geqq 1$ *for* $s \geqq (k-2)2^{k-1} + 5$ *and every integer* n.

Let $p^{w-1} \leqq k < p^w$. Then since $p \nmid (A_k, \cdots, A_1)$, $\theta \leqq w - 1$. By Lemma 5, $N(p^\gamma) \geqq 1$ for $s \geqq p^\gamma + 1$.

For $p > 2$,

$$p^\gamma + 1 \leqq p^w + 1 \leqq pk + 1 \leqq k^2 + 1 < (k-2)2^{k-1} + 5\,.$$

Therefore $N(p^\gamma) \geqq 1$ for $s \geqq (k-2)2^{k-1} + 5$.

For $p = 2$,

$$p^\gamma + 1 \leqq p^{w+1} + 1 \leqq 4k + 1 \leqq (k-2)2^{k-1} + 5\,.$$

Therefore $N(p^\gamma) \geqq 1$ for $s \geqq (k-2)2^{k-1} + 5$.

6. **Completion of the analysis.** By means of the Lemmas 4, 5, and 6 it will be proved that for polynomials of degree 4, 5, 6, and 7 satisfying hypoth-

eses I and II, and for certain polynomials of degree k, $7 < k \leqq 28$, satisfying the same hypotheses, that $N(p^\gamma) \geqq 1$ for $s \geqq (k-2)2^{k-1}+5$ and every integer n.

From this fact and Lemmas 1 and 2, since $M(p^l) \geqq N(p^l)$ we have

$$
\begin{aligned}
\sum_{q/p^l} A(q) &= p^{-l(s-1)}\, M(p^l) \geqq p^{-l(s-1)}\, N(p^l) \\
&= p^{-l(s-1)}\, p^{(l-\gamma)(s-1)}\, N(p^\gamma) \geqq p^{-\gamma(s-1)} \\
&\geqq p^{-\gamma_1(s-1)},
\end{aligned} \tag{7}
$$

where γ_1 is for each value of k the greatest value of γ such that $2^\gamma \leqq (k-2)$ $2^{k-1}+4$.

Also, by Landau,[5] Theorem 8, with $\epsilon = 1/8$

$$
\begin{aligned}
\sum_{q/p^l} A(q) &= 1 + \sum_{\lambda=1}^{l} A(p^\lambda) > 1 - E \sum_{\lambda=1}^{\infty} p^{-9\lambda/8} \\
&= 1 - E(p^{9/8}-1)^{-1}.
\end{aligned} \tag{8}
$$

Finally, it has been proved by Landau[5] that $\mathfrak{S}$ is absolutely convergent for $s \geqq (k-2)2^{k-1}+5$. Hence, by (7), (8), and Lemma 3

$$
\begin{aligned}
\mathfrak{S} &= \lim_{l\to\infty} \sum_{q/p_1^l \cdots p_l^l} A(q) = \lim_{l\to\infty} \prod_{p \leqq p_l} \sum_{q/p^l} A(q) \\
&\geqq \prod_p \max\left(p^{-\gamma_1(s-1)},\, 1 - E(p^{9/8}-1)^{-1}\right) \\
&\geqq \eta > 0 .
\end{aligned}
$$

As indicated in §1, it follows that $r_s(n) > 0$ for $n > C_2$.

7. Polynomials of degree 4.

We prove

THEOREM 1. *If $P(y)$ satisfies hypotheses* I *and* II, *every sufficiently large integer can be represented as a sum of* 21 *values of a polynomial*

$$
P(y) = a_4\begin{bmatrix} y \\ 4 \end{bmatrix} + a_3\begin{bmatrix} y \\ 3 \end{bmatrix} + a_2\begin{bmatrix} y \\ 2 \end{bmatrix} + a_1\begin{bmatrix} y \\ 1 \end{bmatrix},
$$

$a_4, \cdots, a_1$ being integers, $a_4 > 0$. $P(y)$ fails to satisfy hypothesis II *when and only when one or more of the following three sets of simultaneous conditions hold:*

1) $3/(a_4, a_3),\ 3^2 \nmid a_4,\ 2a_3 + 3a_2 + 6a_1 \equiv 2a_4 + 3a_2 \equiv 0 \pmod 9$,

2) $2^3/a_4,\ 2^4 \nmid a_4,\ 2/(a_3, a_2)$,

$a_4 + a_3 + 6a_2 \equiv a_4 + 2a_3 \equiv 0 \pmod{2^4}$,

$3a_4 + 4a_3 + 6a_2 + 12a_1 \equiv 0 \pmod{2^5}$,

3) $2^4/a_4,\ 2/(a_3, a_2),\ a_2 + 2a_1 \equiv a_3 + 2a_2 \equiv 0 \pmod 8$.

First we prove that for $P(y)$ as in the theorem, and $s \geqq 21$, $N(p^\gamma) \geqq 1$.
For $p > k$, this is proved in Lemma 4.
Let $p \leqq k$. Formulas (5) and (6) give for $k = 4$,

$$b_4 = a_4, \qquad b_3 = 6a_4 + 4a_3, \qquad b_2 = 11a_4 + 12a_3 + 12a_2,$$

$$b_1 = 6a_4 + 8a_3 + 12a_2 + 24a_1,$$

$$A_4 = b_4v^4/4!, \qquad A_3 = (4b_4t + b_3)v^3/4!, \qquad A_2 = (6b_4t^2 + 3b_3t + b_2)v^2/4!,$$

$$A_1 = (4b_4t^3 + 3b_3t^2 + 2b_2t + b_1)v/4!, \qquad A_0 = P(t).$$

A. Case $p = 3$.

1. Suppose $3/(a_4, a_3)$, then $3 \nmid v$. By hypothesis I, every value of t is admissible.
   a. Unless $b_1 \equiv 4b_4 + 2b_2 \equiv 0 \pmod 9$, there is a value of t for which $3 \nmid A_1$. Then $\theta = 0$ and $3 \nmid Q_0(0)$. Then by Lemma 5, $N(3) \geqq 1$ for $s \geqq 4$.
   b. If $b_1 \equiv 4b_4 + 2b_2 \equiv 0 \pmod 9$, for every value of t, $A_1 \equiv 4A_4 + 2A_2 \equiv 0 \pmod 3$. If $9 \nmid b_4$, $\theta = 0$ and the polynomial does not satisfy hypothesis II. This case gives exception 1 of the theorem. If $9/b_4$, then $9 \nmid b_3$ and $3 \nmid A_3$. Then $\theta = 1$ and there is a value of x for which $3 \nmid Q_0(x)$. By Lemma 5, $N(3^2) \geqq 1$ for $s \geqq 10$.
2. Suppose $3 \nmid (a_4, a_3)$, then $v \equiv 3, 6 \pmod 9$. $3/(A_4, A_3, A_2)$ and there is a value of t such that $3 \nmid A_1$. Then $\theta = 0$, $3 \nmid Q_0(0)$, and either $3 \nmid Q(0)$ or $3 \nmid Q(1)$ according as $3 \nmid A_0$ or $3/A_0$. By Lemma 5, $N(3) \geqq 1$ for $s \geqq 4$.

B. Case $p = 2$.

1. Suppose $2^3/a_4$, $2/(a_3, a_2)$, then $2 \nmid v$. By hypothesis I, every value of t is admissible.
   Unless $b_1 \equiv b_3 \equiv 0 \pmod{16}$, there is a value of t for which $2 \nmid A_1$. Then $\theta = 0$ and $2 \nmid Q_0(0)$. By Lemma 5, $N(2^2) \geqq 1$ for $s \geqq 5$.
   If $b_1 \equiv b_3 \equiv 0 \pmod{16}$, then $2/A_1$ for all t. There are two cases to be considered.
   a. If $16 \nmid a_4$, then $2 \nmid A_4$, $2/(A_3, A_2, A_1)$. Also $\theta \leqq 2$ and $2 \nmid Q(1)$ or $2 \nmid Q(0)$ according as $2/A_0$ or $2 \nmid A_0$. Then by Lemma 5, $N(2^\gamma) \geqq 1$ for $s \geqq 17$, unless $\theta = 2$ and for every value of t, $A_1 \equiv 4A_4 + 3A_3 + 2A_2 \equiv 0 \pmod 8$, which gives exception 2 of the theorem.
   b. If $16/a_4$, then $2/A_4$, $2 \nmid A_2$, and $\theta = 1$. Then $2 \nmid Q(0)$ or $2 \nmid Q(1)$ according as $2 \nmid A_0$ or $2/A_0$. If for some t, $4 \nmid A_1$, then $2 \nmid Q_0(0)$. Then $N(2^3) \geqq 1$ for $s \geqq 9$. If $4/A_1$ for all values of t, there is no value of t for which $4/A_3$ and the polynomial is under exception 3 of the theorem.
2. Suppose $2/a_4$ but not all the conditions of Case B1 hold. Then $v \equiv 2 \pmod 4$ and $2/(A_4, A_3)$.
   a. Suppose $2a_3 \equiv a_4 + 2a_2 \equiv 0 \pmod 4$ does not hold.
   If $4/a_4$, then $2/A_2$ and there is a value of t for which $2 \nmid A_1$. Then

$\theta = 0$ and $2 \nmid Q_0(0)$. $2 \nmid Q(0)$ or $2 \nmid Q(1)$ according as $2 \nmid A_0$ or $2/A_0$. Then $N(2^2) \geqq 1$ for $s \geqq 5$.

If $4 \nmid A_4$, then $2 \nmid A_2$. By hypothesis I, there is a value of t for which $A_1 \equiv A_0 \equiv 1 \pmod 2$, or $2/A_1$. If A_1 is odd, then $\theta = 0$, $Q_0(0) \equiv Q(0) \equiv 1 \pmod 2$, and $N(2^2) \geqq 1$ for $s \geqq 5$. If A_1 is even, then $2 \nmid Q(1)$ or $2 \nmid Q(0)$ according as $2/A_0$ or $2 \nmid A_0$. Then $\theta = 1$ and, since $4/A_3$, either $2/Q_0(0)$ or $2 \nmid Q_0(1)$ according as $4 \nmid A_1$ or $4/A_1$. Then $N(2^3) \geqq 1$ for $s \geqq 9$.

b. Suppose $2a_3 \equiv a_4 + 2a_2 \equiv 0 \pmod 4$. Then $2/A_1$ for all values of t.
If $4 \nmid a_4$, then $2 \nmid A_2$ and $\theta = 1$. $2 \nmid Q(0)$ or $2 \nmid Q(1)$ according as $2 \nmid A_0$ or $2/A_0$. Since $4/A_3$, we have $2 \nmid Q_0(0)$ or $2 \nmid Q_0(1)$, according as $4 \nmid A_1$ or $4/A_1$ and $N(2^3) \geqq 1$ for $s \geqq 9$.

If $4/a_4$, then $2/A_2$. Choose $t \equiv 1 \pmod 2$. Then $2 \nmid Q(0)$.

If $4 \nmid A_1$, then $\theta = 1$, $2 \nmid Q_0(0)$ and $N(2^3) \geqq 1$ for $s \geqq 9$. If $4/A_1$, $\theta \geqq 2$.

If $8 \nmid A_1$, then $2 \nmid Q_0(0)$ and $N(2^4) \geqq 1$ for $s \geqq 17$.

If $8/A_1$, then $4 \nmid A_2$ since $a_4 \not\equiv 0 \pmod 8$. Then $\theta = 2$, and since $8/A_3$, $2 \nmid Q_0(1)$. Then $N(2^4) \geqq 1$ for $s \geqq 17$.

3. Suppose $2 \nmid a_4$, then $v \equiv 4 \pmod 8$ and $2/(A_4, A_3, A_2)$, $2 \nmid A_1$. Then $\theta = 0$ and $2 \nmid Q_0(0)$. $2 \nmid Q(0)$ or $2 \nmid Q(1)$ according as $2 \nmid A_0$ or $2/A_0$. Then $N(2^2) \geqq 1$ for $s \geqq 5$.

Thus, with the three exceptions listed in the theorem $N(p^\gamma) \geqq 1$, for polynomials of degree 4, $s \geqq 21$, every prime p, and every integer n.

Then as in §5, with $\gamma_1 = 4$, $r_s(n) > 0$ for $n > C_2$.

8. **Polynomials of degree** 5. We prove

THEOREM 2. *If $P(y)$ satisfies hypotheses* I *and* II, *every sufficiently large integer can be represented as a sum of* 53 *values of a polynomial*

$$P(y) = a_5\begin{bmatrix} y \\ 5 \end{bmatrix} + a_4\begin{bmatrix} y \\ 4 \end{bmatrix} + a_3\begin{bmatrix} y \\ 3 \end{bmatrix} + a_2\begin{bmatrix} y \\ 2 \end{bmatrix} + a_1\begin{bmatrix} y \\ 1 \end{bmatrix},$$

the a's being integers, $a_5 > 0$. $P(y)$ fails to satisfy hypothesis II *when and only when one or more of the following eight sets of simultaneous conditions hold.*

1) $9 \nmid a_5$, $9 \nmid a_4$, $a_4 + 3a_2 \equiv a_3 + 6a_2 + 3a_1 \equiv 0 \pmod 9$;

2) $9/(a_5, a_4)$, $3/(a_3, a_2)$, $a_5 + 3a_3 \equiv 2a_5 + a_4 + 3a_3 + 3a_2$
$$\equiv 2a_3 + 3a_2 + 6a_1 \equiv 0 \pmod{27} ;$$

3) $2^3/(a_5, a_4)$, $2^4 \nmid a_5$, $a_3 \equiv a_2 + 2a_1 \equiv 0 \pmod 4$;

4) $2^4/(a_5, a_4)$, $4/a_3$, $a_3 + 2a_2 \equiv a_2 + 2a_1 \equiv 0 \pmod 8$;

5) $a_5 \equiv 2a_4 \equiv 16 \pmod{32}$,

$2a_3 + 4a_2 \equiv a_4 + 2a_2 + 4a_1 \equiv 0 \pmod{16}$;

6) $2^5/a_5$, $a_4 \equiv 2a_3 \equiv 4a_2 \equiv 8 \pmod{16}$, $a_4 + a_3 + 6a_2 \equiv 0 \pmod{16}$,

$3a_4 + 4a_3 + 6a_2 + 12a_1 \equiv 2a_3 + 4a_2 \equiv 0 \pmod{32}$;

7) 1. $a_5 \equiv 2$, $a_4 \equiv 0 \pmod 4$, $a_3 \equiv 1$, $a_2 \equiv 0 \pmod 2$,

$a_5 + a_4 + 6a_3 + 2a_2 + 4a_1 \equiv 0 \pmod 8$,

or 2. $a_5 \equiv a_4 \equiv 2a_3 \equiv 2a_2 \equiv 2 \pmod 4$, $2/a_1$,

$3a_4 + 4a_3 + 6a_2 \equiv 0 \pmod 8$;

8) $a_5 \equiv 4$, $a_4 \equiv 0 \pmod 8$, $a_5 + a_4 + 6a_3 + 2a_2 + 12a_1 \equiv 0 \pmod{16}$.

First we prove that for $P(y)$ as in the theorem and $s \geqq 53$, $N(p^\gamma) \geqq 1$.

For $p > k$, this is proved in Lemma 4.

Let $p \leqq k$. From formulas (5) and (6) we can calculate the coefficients $b_5, \cdots, b_1$ and $A_5, \cdots, A_0$. They are similar to those for $k = 4$ in §5, but for brevity, only the numerical coefficients in the expressions for them are indicated here. These are

$$b_5 = 1, \quad b_4 = 10, 5, \quad b_3 = 35, 30, 20, \quad b_2 = 50, 55, 60, 60,$$

$$b_1 = 24, 30, 40, 60, 120,$$

$$A_5 = 1, \quad A_4 = 5, 1, \quad A_3 = 10, 4, 1, \quad A_2 = 10, 6, 3, 1,$$

$$A_1 = 5, 4, 3, 2, 1, \quad A_0 = P(t).$$

A. Case $p = 5$.

1. Suppose $5/a_5$. Then $5 \nmid v$ and $t \equiv 0 \pmod{25}$ is admissible. Then $\theta \leqq 1$ and since not all coefficients of $Q_0(x)$ are divisible by 5, for some value of x, $5 \nmid Q_0(x)$. Then $N(5^\gamma) \geqq 1$ for $s \geqq 26$.
2. Suppose $5 \nmid a_5$, then $5/v$ and $5/(A_5, A_4, A_3, A_2)$, $5 \nmid A_1$. Then $\theta = 0$, $5/Q_0(0)$, and $5 \nmid Q(0)$ or $5 \nmid Q(1)$ according as $5 \nmid A_0$ or $5/A_0$. Then $N(5) \geqq 1$ for $s \geqq 6$.

B. Case $p = 3$.

1. Suppose $3/(a_5, a_4, a_3)$. Then $3 \nmid v$ and every value of t is admissible.
   a. Unless $b_5 \equiv 4b_4 + 2b_2 \equiv b_1 \equiv 0 \pmod 9$, let $t \equiv 0 \pmod 9$. Then $\theta = 0$ and there is a value of x for which $3 \nmid Q_0(x)$. Then by Lemma 5, $N(3) \geqq 1$ for $s \geqq 4$.
   b. Suppose $b_5 \equiv 4b_4 + 2b_2 \equiv b_1 \equiv 0 \pmod 9$. If $b_4 \not\equiv 0 \pmod 9$, then $3 \nmid A_4$ and $\theta = 0$. Then for all values of t, $A_5 \equiv 4A_4 + 2A_2 \equiv A_1 \equiv 0 \pmod 3$, and the polynomial does not satisfy hypothesis II. This is exception 1 of the theorem.
   If $b_4 \equiv 0 \pmod 9$, but $5b_5 + 3b_3 \equiv 4b_4 + 2b_2 \equiv b_1 \equiv 0 \pmod{27}$ does not hold, choose $t \equiv 0 \pmod{27}$. Then, since $3 \nmid b_3$, $\theta = 1$ and there is a value of x for which $3 \nmid Q_0(x)$. Then $N(3^2) \geqq 1$ for $s \geqq 10$. If $b_4 \equiv 0 \pmod 9$ and $5b_5 + 3b_3 \equiv 4b_4 + 2b_2 \equiv b_1 \equiv 0 \pmod{27}$, then $\theta = 1$. Since, for all t, $5A_5 + 3A_3 \equiv 4A_4 + 2A_2 \equiv A_1 \equiv 0 \pmod 9$, the poly-

nomial does not satisfy hypothesis II. This is exception 2 of the theorem.

2. Suppose $3 \nmid (a_5, a_4, a_3)$, then $v \equiv 3, 6 \pmod 9$ and $3/(A_5, \cdots, A_2)$. $3/A_1$ for every value of t implies $3/(a_5, a_4, a_3)$. Therefore t can be chosen so that $3 \nmid A_1$. Then $\theta = 0$, and $3 \nmid Q_0(0)$ and $3 \nmid Q(0)$ or $3 \nmid Q(1)$ according as $3 \nmid A_0$ or $3/A_0$. Then $N(3) \geqq 1$ for $s \geqq 4$.

C. Case $p = 2$.

1. Suppose $2^3/(a_5, a_4)$, $2/(a_3, a_2)$, then $2 \nmid v$ and every value of t is admissible.
   a. Unless $a_3 \equiv a_2 + 2a_1 \equiv 0 \pmod 4$, there is a value of t such that $2 \nmid A_1$. Then $\theta = 0$, $2 \nmid Q_0(0)$, and $N(2^2) \geqq 1$ for $s \geqq 5$.
   b. If $a_3 \equiv a_2 + 2a_1 \equiv 0 \pmod 4$ and $2^4 \nmid a_5$, the polynomial does not satisfy hypothesis II. This is exception 3 of the theorem.
   c. Suppose $a_3 \equiv a_2 + 2a_1 \equiv 0 \pmod 4$ and $2^4/a_5$.
   If $2^4 \nmid a_4$, then $2 \nmid A_2$ and $\theta = 1$. If there is a value of t for which $4 \nmid A_1$, then $2 \nmid Q_0(0)$ and $N(2^3) \geqq 1$ for $s \geqq 9$. If $4/A_1$ for all t, then $2^3 \nmid a_3$, and $5A_5 + 3A_3 + 2A_2 \equiv 0 \pmod 4$ for all t. The polynomial does not satisfy hypothesis II. This is exception 4 of the theorem.
   If $2^4/a_4$, then $2/A_2$. If for some value of t, $4 \nmid A_1$, then $\theta = 1$ and $2 \nmid Q_0(0)$. Then $N(2^3) \geqq 1$ for $s \geqq 9$. If $4/A_1$ for all t, $2^3 \nmid a_3$. If $2^5 \nmid a_5$, $\theta = 1$ and $4/(5A_5 + 3A_3)$ for all t and the polynomial does not satisfy the hypothesis II. This is exception 5 of the theorem. If $2^5/a_5$, then $\theta = 2$. If for some value of t, $8 \nmid A_1$, $N(2^4) \geqq 1$ for $s \geqq 17$. If $8/A_1$ for all t, and $a_3 + 2a_2 \equiv 8 \pmod{16}$, then $8 \nmid (5A_5 + 4A_4 + 3A_3 + 2A_2)$. Then $N(2^4) \geqq 1$ for $s \geqq 17$. If $a_3 + 2a_2 \equiv 0 \pmod{16}$, the polynomial does not satisfy hypothesis II. This is exception 6 of the theorem.

2. Suppose $2/(a_5, a_4)$ but not all the conditions of C1 hold. Then $v \equiv 2 \pmod 4$ and $2/(A_5, A_4, A_3)$.
   a. Unless $a_3 \equiv a_2 \equiv 0 \pmod 2$, there is a value of t such that $2 \nmid (A_2 + A_1)$. If for the same value of t, $2 \nmid A_1$, then $\theta = 0$ and $2 \nmid Q_0(0)$. $2 \nmid Q(0)$ or $2 \nmid Q(1)$ according as $2 \nmid A_0$ or $2/A_0$. Then $N(2^2) \geqq 1$ for $s \geqq 5$. If for every value of t for which $2 \nmid (A_2 + A_1)$, $2/A_1$, then $2 \nmid A_2$. Then $\theta = 1$ and $2 \nmid Q(0)$ or $2 \nmid Q(1)$ according as $2 \nmid A_0$ or $2/A_0$. If $4 \nmid A_1$, we have $2 \nmid Q_0(0)$, and $N(2^3) \geqq 1$ for $s \geqq 9$. If $4/A_1$ and $4 \nmid a_5$, then $2 \nmid Q_0(1)$ and $N(2^3) \geqq 1$ for $s \geqq 9$. If $4/A_1$ and $4/a_5$, then $4 \nmid A_3$ and there is a value of t for which $2/A_2$. If for this value $2 \nmid A_0$, then $\theta = 1$ and $2 \nmid Q_0(1)$ or $2 \nmid Q_0(0)$ according as $4/A_1$ or $4 \nmid A_1$. $2 \nmid Q(0)$ and $N(2^2) \geqq 1$ for $s \geqq 5$. If for this value of t we have $2/A_0$, then the polynomial does not satisfy hypothesis II. This is exception 7 of the theorem.
   b. If $a_3 \equiv a_2 \equiv 0 \pmod 2$, then $2/(A_2 + A_1)$. Choose $t \equiv 1 \pmod 2$, then $2 \nmid Q(0)$.
   If $a_5 \not\equiv a_4 \pmod 4$, then $2 \nmid A_1$, $\theta = 0$, and $2 \nmid Q_0(0)$. Then $N(2^2) \geqq 1$ for $s \geqq 5$.
   If $a_5 \equiv a_4 \pmod 4$, $\theta \geqq 1$. Then if $4 \nmid A_1$, $\theta = 1$ and $N(2^3) \geqq 1$ for

$s \geqq 9$. If $4/A_1$ and $4 \nmid a_5$, then $4 \nmid A_3$, $\theta = 1$ and $N(2^3) \geqq 1$ for $s \geqq 9$. If $4/A_1$ and $4/a_5$, $\theta = 2$. Then if $8 \nmid A_1$, $N(2^4) \geqq 1$ for $s \geqq 17$. If $8/A_1$ for all values of $t \equiv 1 \pmod 2$, unless $8 \nmid a_5$ and $8/a_4$, $2 \nmid Q_0(1)$ and $N(2^4) \geqq 1$ for $s \geqq 17$. If $8 \nmid a_5$, $8/a_4$ the polynomial does not satisfy hypothesis II. This is exception 8 of the theorem.

3. Suppose $2 \nmid (a_5, a_4)$. Then $v \equiv 4 \pmod 8$ and $2/(A_5, \cdots, A_2)$. Since $2/A_1$ for all values of t implies $2/(a_5, a_4)$, there is a value of t for which $2 \nmid A_1$. Then $\theta = 0$, $2 \nmid Q_0(0)$ and $2 \nmid Q(0)$ or $2 \nmid Q(1)$ according as $2 \nmid A_0$ or $2/A_0$. Then $N(2^2) \geqq 1$ for $s \geqq 5$.

Thus, with the eight exceptions listed in the theorem, $N(p^\gamma) \geqq 1$, for polynomials of degree 5, $s \geqq 53$, every prime p, and every integer n.

Then, as in §5, with $\gamma_1 = 5$, $r_s(n) > 0$ for $n > C_2$.

9. Polynomials of degree 6. We prove

THEOREM 3. *If $P(y)$ satisfies hypotheses* I *and* II, *every sufficiently large integer can be expressed as a sum of* 133 *values of a polynomial*

$$P(y) = a_6\begin{bmatrix} y \\ 6 \end{bmatrix} + a_5\begin{bmatrix} y \\ 5 \end{bmatrix} + a_4\begin{bmatrix} y \\ 4 \end{bmatrix} + a_3\begin{bmatrix} y \\ 3 \end{bmatrix} + a_2\begin{bmatrix} y \\ 2 \end{bmatrix} + a_1\begin{bmatrix} y \\ 1 \end{bmatrix},$$

$a_6, \ldots, a_1$ *being integers*, $a_6 > 0$.

It is assumed from the beginning that $P(y)$ satisfies hypothesis II. We prove first that for a polynomial as in the theorem and $s \geqq 133$, $N(p^\gamma) \geqq 1$.

For $p > k$ this is proved in Lemma 4.

Let $p \leqq k$. Formulas (5) and (6) give for $k = 6$

$$b_6 = 1, \quad b_5 = 15, 6, \quad b_4 = 85, 60, 30, \quad b_3 = 225, 210, 180, 120,$$

$$b_2 = 274, 300, 330, 360, 360, \quad b_1 = 120, 144, 180, 240, 360, 720,$$

$$A_6 = 1, \quad A_5 = 6, 1, \quad A_4 = 15, 5, 1, \quad A_3 = 20, 10, 4, 1,$$

$$A_2 = 15, 10, 6, 3, 1, \quad A_1 = 6, 5, 4, 3, 2, 1, \quad A_0 = P(t).$$

A. Case $p = 5$.

1. Suppose $5/(a_6, a_5)$, then $5 \nmid v$ and every value of t is admissible. By hypothesis I, at least one of $A_6, \cdots, A_1$ is not divisible by 5. Then $\theta \leqq 1$. Choose t such that for some value of x, $5 \nmid Q_0(x)$. Then $N(5^\gamma) \geqq 1$ for $s \geqq 26$.
2. Suppose $5 \nmid (a_6, a_5)$, then $5/v$ and $5/(A_6, \cdots, A_2)$. There is a value of t such that $5 \nmid A_1$, since $5/A_1$ for all t implies $5/(a_6, a_5)$. Therefore $\theta = 0$, $5 \nmid Q_0(0)$, and $5 \nmid Q(0)$ or $5 \nmid Q(1)$ according as $5 \nmid A_0$ or $5/A_0$. Then $N(5) \geqq 1$ for $s \geqq 6$.

B. Case $p = 3$.

1. Suppose $3^2/a_5$, $3/(a_5, a_4, a_3)$, then $3 \nmid v$. Then as in A1 above, $\theta \leqq 1$ and $N(3^\gamma) \geqq 1$ for $s \geqq 10$.
2. Suppose not all conditions of B1 hold. Then $v \equiv 3, 6 \pmod 9$ and $3/(A_6, \cdots, A_3)$.

a. If $3 \mid a_6$, then $3 \mid A_2$. If there is a value of t for which $3 \nmid A_1$, $\theta = 0$ and $3 \nmid Q_0(0)$. $3 \nmid Q(0)$ or $3 \nmid Q(1)$ according as $3 \nmid A_0$ or $3 \mid A_0$. Then $N(3) \geqq 1$ for $s \geqq 4$. If $3 \mid A_1$ for all t, then $3 \mid (a_5, a_4, a_3)$. Then since $3^2 \nmid a_6$, $3^2 \nmid A_2$, and $\theta = 1$. Choose t so that $3 \nmid A_0$. Then $3 \nmid Q(0)$ and $3 \nmid Q_0(0)$ or $3 \nmid Q_0(1)$ according as $3^2 \nmid A_1$ or $3^2 \mid A_1$. Then $N(3^2) \geqq 1$ for $s \geqq 10$.

b. If $3 \nmid a_6$, then $3 \nmid A_2$ and $\theta = 0$. Then there is a value of x for which $3 \nmid Q(x)$. $3 \nmid Q_0(0)$ or $3 \nmid Q_0(1)$ according as $3 \nmid A_1$, or $3 \mid A_1$. Then $N(3) \geqq 1$ for $s \geqq 4$.

C. Case $p = 2$.

1. Suppose $2^4 \mid a_6$, $2^3 \mid (a_5, a_4)$, $2 \mid (a_3, a_2)$. Then $2 \nmid v$. As in A1 $\theta \leqq 2$ and $N(2^\gamma) \geqq 1$ for $s \geqq 17$.

2. Suppose $2 \mid (a_6, a_5, a_4)$ and not all the conditions of C1 hold. Then $v \equiv 2 \pmod 4$ and $2 \mid) A_6, \cdots, A_4)$.

   a. If $2^2 \nmid a_6$, then $2 \nmid A_3$ and $\theta = 0$. Choose t admissible and such that for some value of x, $2 \nmid Q_0(x)$. Then $N(2^2) \geqq 1$ for $s \geqq 5$.

   b. If $2^2 \mid a_6$, $a_3 \not\equiv a_2$, $a_3 \equiv 0 \pmod 2$, then for every value of t, $2 \nmid (A_2 + A_1)$. Choose t such that for some value of x, $2 \nmid Q_0(x)$. $2 \nmid Q(0)$ or $2 \nmid Q(1)$ according as $2 \nmid A_0$ or $2 \mid A_0$. Then $\theta \leqq 1$ and $N(2^\gamma) \geqq 1$ for $s \geqq 9$. If $a_3 \equiv 1 \pmod 2$, choose $t \equiv 1 \pmod 2$. Then $2 \nmid (A_2 + A_1)$. Since $t \equiv 0 \pmod 2$ is not admissible, there is a value of x for which $2 \nmid Q_0(x)$. As above, $N(2^\gamma) \geqq 1$ for $s \geqq 9$.

   c. Let $2^2 \mid a_6$, $a_3 \equiv a_2$, $a_3 \equiv 1 \pmod 2$. If for $t \equiv 0 \pmod 2$ there is a value of x for which $2 \nmid Q_0(x)$, then $\theta \leqq 1$ and $2 \nmid Q(0)$ or $2 \nmid Q(1)$ according as $2 \nmid A_0$ or $2 \mid A_0$. Then $N(2^\gamma) \geqq 1$ for $s \geqq 9$. Otherwise choose $t \equiv 1 \pmod 2$. Since, if $a_3 \equiv 0 \pmod 2$, $t \equiv 1 \pmod 2$ is the only admissible value of t, we have now to consider the case $2^2 \mid a_3$ and $a_3 \equiv a_2 \pmod 2$, $t \equiv 1 \pmod 2$. If $a_5 \not\equiv a_4 \pmod 4$, then $2 \nmid A_1$. $\theta = 0$, $2 \nmid Q_0(0)$, $2 \nmid Q(0)$. Then $N(2^\gamma) \geqq 1$ for $s \geqq 5$. If $a_5 \equiv a_4 \pmod 4$ and $a_5 \not\equiv 0 \pmod 4$, then $\theta \leqq 5$, since for $t \equiv 1 \pmod 2$, $2^6 \nmid (2A_6, A_5)$. Then $N(2^\gamma) \geqq 1$ for $s \geqq 129$. If $a_5 \equiv a_4 \pmod 4$ and $a_5 \equiv 0 \pmod 4$, $\theta \leqq 5$. For if $\theta = 6$, for $t \equiv 1 \pmod 2$,

$$2^6 / 6A_6 \text{ implies } \qquad a_6 \equiv 0 \pmod{2^3},$$

$$2^6 / 5A_5 \text{ implies } 5a_6 + 6a_5 \equiv 0 \pmod{2^5},$$

$$2^6 / 4A_4 \text{ implies } 3\,a_6 + 2a_5 + 2a_4 \equiv 0 \pmod{2^4}.$$

   These together imply one of

   1) $16 \mid a_6$, $8 \mid (a_5, a_4)$. In this case the polynomial belongs in case C1.

   2) $16 \nmid a_6$, $8 \nmid a_5$, $8 \mid a_4$. Then if $2 \mid a_3$, $8 \nmid 2A_2$; and if $2 \nmid a_3$, $8 \nmid A_3$. Thus $\theta \leqq 5$, and $N(2^\gamma) \geqq 1$ for $s \geqq 129$.

3. Suppose $2 \nmid (a_6, a_5, a_4)$, then $v \equiv 4 \pmod 8$. Then $2 \mid (A_6, \cdots, A_2)$. Unless $a_6 \equiv a_5 \equiv 1$, $a_4 \equiv 0 \pmod 2$, there is a value of t for which $2 \mid A_1$. Then $\theta = 0$, $2 \nmid Q_0(0)$, and $2 \nmid Q(0)$ or $2 \nmid Q(1)$ according as $2 \nmid A_0$ or $2 \mid A_0$. Then $N(2^2) \geqq 1$ for $s \geqq 5$.

If $a_6 \equiv a_5 \equiv 1$, $a_4 \equiv 0 \pmod 2$, since $8 \nmid A_3$, $\theta \leqq 2$. Choose t so that $2 \nmid A_0$ and for some value of x, $2 \nmid Q_0(x)$. Then $N(2^\gamma) \geqq 1$ for $s \geqq 17$.

Thus, for $s \geqq 133$, $N(p^\gamma) \geqq 1$, for polynomials of degree 6 satisfying hypotheses I and II.

Then as in §5, with $\gamma_1 = 5$, $r_s(n) > 0$ for $n > C_2$.

10. Polynomials of degree 7. We prove

THEOREM 4. *If $P(y)$ satisfies hypotheses* I *and* II, *every sufficiently large integer can be expressed as a sum of* 325 *values of a polynomial*

$$P(y) = a_7 \begin{bmatrix} y \\ 7 \end{bmatrix} + \cdots + a_1 \begin{bmatrix} y \\ 1 \end{bmatrix},$$

$a_7, \cdots, a_1$ *being integers*, $a_7 > 0$.

It is assumed from the beginning that $P(y)$ satisfies hypothesis II. We prove first that for a polynomial as in the theorem and $s \geqq 325$, $N(p^\gamma) \geqq 1$.

For $p > k$ this is proved in Lemma 4.

Let $p \leqq k$. Formulas (5) and (6) give for $k = 7$

$$\begin{aligned}
b_7 &= 1, \ b_6 = 21, 7, \ b_5 = 175, 105, 42, \\
b_4 &= 735, 595, 420, 210, \ b_3 = 1624, 1575, 1470, 1260, 840, \\
b_2 &= 1764, 1918, 2100, 2310, 2520, 2520, \\
b_1 &= 720, 840, 1008, 1260, 1680, 2520, 5040, \\
A_7 &= 1, \ A_6 = 7, 1, \ A_5 = 21, 6, 1, \ A_4 = 35, 15, 5, 1, \\
A_3 &= 35, 20, 10, 4, 1, \ A_2 = 21, 15, 10, 6, 3, 1, \\
A_1 &= 7, 6, 5, 4, 3, 2, 1, \ A_0 = P(t).
\end{aligned}$$

A. Case $p = 7$.

1. Suppose $7 / a_7$, then $7 \nmid v$ and every value of t is admissible. At least one of $A_7, \cdots, A_1$ is not divisible by 7. Then $\theta \leqq 1$ and $N(7^\gamma) \geqq 1$ for $s \geqq 50$.
2. Suppose $7 \nmid a_7$. Then $v \equiv 7, 14, \cdots, 42 \pmod{49}$, and $7 / (A_7, \cdots, A_2)$ but $7 \nmid A_1$. Thus $\theta = 0$, $7 \nmid Q_0(0)$ and $7 \nmid Q(0)$ or $7 \nmid Q(1)$ according as $7 \nmid A_0$ or $7 / A_0$. Then $N(7) \geqq 1$ for $s \geqq 8$.

B. Case $p = 5$.

1. Suppose $5 / (a_7, a_6, a_5)$, then $5 \nmid v$. As in A1, $\theta \leqq 1$. Then $N(5^\gamma) \geqq 1$ for $s \geqq 26$.
2. Suppose $5 \nmid (a_7, a_5, a_5)$. Then $v \equiv 5, 10, 15, 20 \pmod{25}$, $5 / (A_7, \cdots, A_2)$ and there is a value of t for which $5 \nmid A_1$. Thus $\theta = 0$, $5 \nmid Q_0(0)$, and $5 \nmid Q(0)$ or $5 \nmid Q(1)$ according as $5 \nmid A_0$ or $5 / A_0$. Then $N(5) \geqq 1$ for $s \geqq 6$.

C. Case $p = 3$.

1. Suppose $3^2 / (a_7, a_6)$, $3 / (a_5, a_4, a_3)$, then $3 \nmid v$. As in A1 $\theta \leqq 1$, and $N(3^\gamma) \geqq 1$ for $s \geqq 10$.

2. Suppose not all conditions of case C1 hold. Then $v \equiv 3, 6 \pmod 9$ and $3 \mid (A_7, \cdots, A_3)$. Unless $3 \mid (a_7, a_6)$, there is a value of t for which $3 \nmid A_2$. Then $\theta = 0$ and there is a value of x for which $3 \nmid Q(x)$. $3 \nmid Q_0(0)$ or $3 \nmid Q_0(1)$ according as $3 \nmid A_1$ or $3 \mid A_1$. Then $N(3) \geqq 1$ for $s \geqq 4$.
   If $3 \mid (a_7, a_6)$, then $3 \mid A_2$ for all t. Unless $3 \mid (a_5, a_4, a_3)$, there is a value of t for which $3 \nmid A_1$. Then $\theta = 0$, $3 \nmid Q_0(0)$, and there is a value of x for which $3 \nmid Q(x)$. Then $N(3) \geqq 1$ for $s \geqq 4$.
   If $3 \mid (a_7, a_6, a_5, a_4, a_3)$, then $3^2 \nmid (a_7, a_6)$. Choose t so that $3 \nmid Q(0)$, and that there is a value of x for which $3 \nmid Q_0(x)$. If $9 \mid a_7$, then $9 \nmid a_6$ and $3^2 \nmid A_2$. Then $\theta \leqq 2$ and $N(3^\gamma) \geqq 1$ for $s \geqq 10$. If $9 \nmid a_7$, then $\theta \leqq 4$ since $3^5 \nmid A_5$, and $N(3^\gamma) \geqq 1$ for $s \geqq 244$.

D. Case $p = 2$.

1. Suppose $2^4 \mid (a_7, a_6)$, $2^3 \mid (a_5, a_4)$, $2 \mid (a_3, a_2)$, then $2 \nmid v$. As in A1 $\theta \leqq 2$ and $N(2^\gamma) \geqq 1$ for $s \geqq 17$.
2. Suppose $2 \mid (a_7, \cdots, a_4)$, and not all the conditions of case D1 hold, then $v \equiv 2 \pmod 4$, and $2 \mid (A_7, \cdots, A_4)$. If $2 \nmid (a_3, a_2)$, there is at least one value of t for which $A_3 + A_2 + A_1 \not\equiv 0 \pmod 2$. Then $2 \nmid Q(0)$ or $2 \nmid Q(1)$ according as $2 \nmid A_0$ or $2 \mid A_0$. If for such a value of t there is a value of x for which $2 \nmid Q_0(x)$, since $2 \nmid (A_3, A_2, A_1)$, $\theta \geqq 1$. Then $N(2^\gamma) \geqq 1$ for $s \geqq 9$. Otherwise, choose t such that $A_3 + A_2 + A_1 \equiv 0 \pmod 2$, $2 \nmid A_0$ and there is a value of x for which $2 \nmid Q_0(x)$. In this case $2 \nmid a_3$. Then if $2 \mid a_2$, the only value of t for which $A_3 + A_2 + A_1 \equiv 0 \pmod 2$ is $t \equiv 0 \pmod 2$ and this is not admissible. Thus $2 \nmid a_2$. We have left to consider then $a_3 \equiv a_2 \pmod 2$ with $t \equiv 1 \pmod 2$, the only admissible value of t. Then $2 \nmid Q(0)$ and there is a value of x for which $2 \nmid Q_0(x)$. $\theta \leqq 6$, for if $\theta = 7$, $A_7 \equiv 2A_6 \equiv A_5 \equiv 4A_4 \equiv 0 \pmod{2^7}$, and these congruences imply $2^4 \mid (a_7, a_6)$, $2^3 \mid (a_5, a_4)$. If $a_3 \equiv a_2 \equiv 1 \pmod 2$, $2^3 \nmid A_3$, and if $a_3 \equiv a_2 \equiv 0 \pmod 2$, the polynomial belongs in case D1. Thus $\theta \leqq 6$ and $N(2^\gamma) \geqq 1$ for $s \geqq 257$.
3. Suppose $2 \nmid (a_7, \cdots, a_4)$, then $v \equiv 4 \pmod 8$ and $2 \mid (A_7, \cdots, A_2)$. Unless $a_7 + a_6 + a_5 \equiv a_4 \equiv 0 \pmod 2$, there is a value of t for which $2 \nmid A_1$. Then $\theta = 0$, $2 \nmid Q_0(0)$ and $2 \nmid Q(0)$ or $2 \nmid Q(1)$. Then $N(2^\gamma) \geqq 1$ for $s \geqq 5$. If $a_7 + a_6 + a_5 \equiv a_4 \equiv 0 \pmod 2$, one of a_7, a_6, a_5 is even and the other two odd. Choose t so that $2 \nmid A_0$ and so that there is a value of x for which $2 \nmid Q_0(x)$. If $2 \mid a_7$, then $8 \nmid A_3$. Then $\theta \leqq 2$ and $N(2^\gamma) \geqq 1$ for $s \geqq 17$. If $2 \mid a_6$ and $t \equiv 1 \pmod 2$, then $8 \nmid A_3$. If $t \equiv 0 \pmod 2$, $2^7 \nmid A_5$. In either case $\theta \leqq 6$ and therefore $N(2^\gamma) \geqq 1$ for $s \geqq 257$. Similarly, if $2 \mid a_5$, $\theta \leqq 6$ and $N(2^\gamma) \geqq 1$ for $s \geqq 257$.

Thus, for $s \geqq 325$, $N(p^\gamma) \geqq 1$, for polynomials of degree 7 satisfying hypotheses I and II.

Then as in §5, with $\gamma_1 = 9$, $r_s(n) > 0$ for $n > C_2$.

11. **Polynomials of degree** k, $7 < k \leqq 28$. We prove

THEOREM 5. *If $P(y)$ satisfies hypotheses* I *and* II *and if for each prime $p \leqq k$ one of the following conditions holds*

(1) *For some admissible value of t with which $P(y)$ satisfies hypothesis* II, $p \nmid (A_k, \cdots, A_1)$,

(2) *For some admissible value of t, with which $P(y)$ satisfies hypothesis* II, $\gamma \leqq \gamma_0$, *where γ_0 is the greatest value of γ such that $p^\gamma \leqq (k-2)2^{k-1} + 5$,*
then every sufficiently large integer can be expressed as a sum of $s \geqq (k-2)2^{k-1} + 5$ values of a polynomial

$$P(y) = a_k \begin{bmatrix} y \\ k \end{bmatrix} + \cdots + a_1 \begin{bmatrix} y \\ 1 \end{bmatrix}, \qquad 7 < k \leqq 28,$$

$a_k, \cdots, a_1$ being integers, $a_k > 0$.

We prove first that $N(p^\gamma) \geqq 1$.

For $p > k$ this is proved by Lemma 4.

Let $p \leqq k$. If condition (1) holds, $N(p^\gamma) \geqq 1$ by Lemma 6. If condition (2) holds, $N(p^\gamma) \geqq 1$ by Lemma 5.

Then as in §5, $r_s(n) > 0$ for $n > C_2$.

UNIVERSITY OF CHICAGO.

CONNECTIONS BETWEEN DIFFERENTIAL GEOMETRY AND TOPOLOGY

I. SIMPLY CONNECTED SURFACES

By Sumner Byron Myers

Introduction. In this paper is presented a theory of new connections between differential geometry and topology. With an arbitrary point A of a complete analytic Riemannian surface S is associated a locus of "minimum points with respect to A". A point M on a geodesic ray g issuing from A is said to be a "minimum point with respect to A on g" if M is the last point on g such that AM furnishes an absolute minimum to the arc length of curves on S joining A to M. The locus of such points with respect to A is proved to be a linear graph m. If S is simply connected, m is a tree when S is closed and a set of infinite trees when S is open. In a later paper, it will be proved that in the general case of a closed multiply connected S, m is a linear graph whose cyclomatic number is equal to the connectivity number modulo 2 of S. The surface S is thus reduced to a single 2-cell σ with m as its singular boundary; σ is simply covered (except at A) by the geodesic rays through A cut off at their intersections with m and hence can be represented by the geodesic polar coördinate system with A as pole. This solves completely the hitherto vaguely answered question as to how long the geodesics through a point A of a surface form a field.

This paper is restricted mainly to simply connected surfaces, hence surfaces homeomorphic to the plane or the sphere. The end points of the branches of the tree (or trees) which forms the locus of minimum points with respect to an arbitrary point A on S are shown to be conjugate to A and to be cusps of the locus of first conjugate points to A. The order of a point M of m as a vertex of m (i.e., the number of arcs of m issuing from M) is proved to be equal to the number of geodesics joining M to A on which M is a minimum point with respect to A.

In order to prove that in the case of a closed analytic simply connected surface the locus m has a finite number of end points, it is necessary to study the locus of first conjugate points to A. This is done in §3. §1 recalls the definition of a complete analytic Riemannian surface, while in §2 the machinery is set up for finding the conjugate point locus. In §4 the minimum point locus is studied. In §5 examples are given of the minimum point locus on a few simple surfaces, while in §6 problems arising from the methods and results of this paper are suggested and possibilities of generalization discussed.

Received April 17, 1935; presented to the American Mathematical Society December 28, 1934. The author is National Research Fellow.

1. Complete surfaces. We are concerned here with *analytic Riemannian surfaces.*[1] These are, in the first place, homogeneous topological surfaces; that is, Hausdorff spaces in which the neighborhood of every point is homeomorphic to the interior of the unit circle in the 2-dimensional euclidean plane. These neighborhoods are provided with euclidean coördinate systems (x_1, x_2) in such a way that in the region of intersection of two different neighborhoods one coördinate system can be obtained from the other by means of an analytic transformation with non-vanishing functional determinant. Any coördinate system which can be obtained from one of these euclidean coördinate systems by means of an analytic transformation with non-vanishing jacobian will be called *admissible.* Finally, every admissible coördinate system is provided with a positive definite symmetric quadratic form $\sum_{i,j=1}^{n} g_{ij}u_iu_j$ in which the functions $g_{ij}(x_1, x_2)$ are real and analytic, and such that if (x_1, x_2) and (x_1', x_2') are the coördinates in two overlapping coördinate systems and $\Sigma g_{ij}u_iu_j$ and $\Sigma g_{ij}'u_iu_j$ the corresponding quadratic forms, then

$$\Sigma g_{ij}dx_idx_j = \Sigma g_{ij}'dx_i'dx_j' .$$

We define arc length on such a surface in the usual manner,

$$s = \int \sqrt{g_{ij}dx_idx_j} .$$

No assumption is made concerning the imbedding of the surface S in any euclidean space, since we are interested only in intrinsic properties of the surface. The words *closed* and *open* when applied to S are synonymous respectively with *compact* and *not compact* in the usual senses.

For convenience in studying properties of the surface S in the large as well as local differential geometry properties, we restrict ourselves to *complete* analytic Riemannian surfaces (or, for brevity, *complete surfaces*) as defined by Hopf and Rinow.[2] This notion of completeness may be defined in either of the following two useful ways:

(1) Every geodesic ray on S can be continued to infinite length.

(2) Every infinite bounded set of points on S has a limit point on S. (Boundedness on S is based on a definition which assigns as distance between two points the lower limit of the lengths of arcs on S joining the two points).

For a proof of the equivalence of these two definitions of completeness as well as a justification of the restriction to complete surfaces, see Hopf and Rinow, loc. cit. On a complete surface any pair of points can be joined by a curve of

[1] This definition is taken from Hopf and Rinow, *Ueber den Begriff der vollständigen differentialgeometrischen Fläche*, Commentarii Mathematici Helvetici, vol. 3 (1931), p. 209. Compare the surfaces used by Veblen and Whitehead in their *The Foundations of Differential Geometry*, and those of Morse in his Colloquium Lectures, *The Calculus of Variations in the Large*, pp. 107–108.

[2] Loc. cit.

shortest length, which is a geodesic. We note that according to the second definition of completeness every closed surface is complete.

Rinow[3] has proved a uniqueness theorem which shows that a complete simply connected continuation[4] of a Riemannian element E is unique, i.e., any two complete simply connected continuations of the same element are isometric. Thus a surface homeomorphic to the sphere cannot have the same differential geometry as a complete surface homeomorphic to the plane, and one of the objects of this paper is to obtain relations between the topological differences and the differential geometry differences for these two types of surfaces.

2. The function $f(r, \theta)$. Suppose we have a complete surface S containing the element E about the point A. It is very convenient to use geodesic polar coördinates (r, θ) with A as pole, so that the Riemannian metric becomes

$$ds^2 = dr^2 + [f(r, \theta)]^2 \, d\theta^2 . \tag{2.1}$$

The function $f(r, \theta)$ has the following properties:[5]

$$\text{(a)} \quad f(0, \theta) \equiv 0 .$$

$$\text{(b)} \quad f_r(0, \theta) \equiv 1 . \tag{2.2}$$

$$\text{(c)} \quad \text{For } 0 \leqq r < \epsilon ,$$

$$[f(r, \theta)]^2 = r^2 + \alpha r^4 + r^5 (\beta \cos \theta + \gamma \sin \theta) + \cdots ,$$

where α, β, γ, etc., are constants.

$$\text{(d)} \quad f(r, \theta) = f(r, \theta + 2\pi) .$$

On the geodesic ray g: $\theta = \theta_0$ the zeros $r \neq 0$ of f give the points on g conjugate to A. Furthermore, the function $f(r, \theta_0)$ satisfies the differential equation

$$f'' + K(r, \theta_0) f = 0 \qquad [f'' \equiv f_{rr}(r, \theta_0)] , \tag{2.3}$$

where $K(r, \theta)$ is the Gaussian curvature at the point (r, θ).

These polar coördinates (r, θ) with A as pole will be useful not only in the neighborhood of A itself, but throughout S. Of course a point P on S will in general have many coördinate pairs (r, θ) belonging to it. Every point on S has at least one pair (r, θ) belonging to it, for every two points on S can be joined by a geodesic.

As Rinow has shown,[6] if E is continuable to a complete surface, the function

[3] *Ueber Zusammenhänge zwischen der Differentialgeometrie im Grossen und im Kleinen*, Math. Zeitschrift, vol. 35 (1932), p. 514. The present author has generalized this theorem to n dimensions. See Myers, *Riemannian manifolds in the large*, this Journal, vol. 1 (1935), pp. 39–49.

[4] A surface S is said to be a *continuation* of a Riemannian element E if S contains an element isometric to E.

[5] See Blaschke, *Vorlesungen über Differentialgeometrie*, I, 3rd edition (1930), p. 152, §70.

[6] See Rinow, paper cited in footnote 3, p. 520.

$f(r, \theta)$ must be an analytic function of r for $\theta =$ constant and all $r > 0$. A much stronger result is the following

THEOREM 1. *If the element E is continuable to a complete surface S, $f(r, \theta)$ as a function of two variables can be continued analytically for all θ and for all $r > 0$. Furthermore, the function $K(r, \theta)$, which equals $-f_{rr}/f$, has the same property.*

First we must state a fundamental lemma.

LEMMA 1. *Points on geodesics of S through an arbitrary point A depend analytically on initial directions and lengths, i.e., if (x, y) forms a local admissible coördinate system on S and if the point (x_0, y_0) lies on a geodesic arc whose initial point is A and whose initial direction is θ_0, and whose length is r_0, then $x = x(r, \theta)$, $y = y(r, \theta)$, where these functions are analytic in the neighborhood of (r_0, θ_0) and remain analytic as long as (r, θ) gives a value of (x, y) in the range of the coördinate system (x, y).*

This is a well-known property of geodesics.

Now we proceed to a proof of Theorem 1.

Let g: $\theta = \theta_0$ be an arbitrary geodesic ray issuing from A, and let B: (r_0, θ_0) be a point just before the first conjugate point to A on g. Then the geodesics through A neighboring g form a field containing B, and the coördinates (r, θ) form an admissible coördinate system in this field. Hence $f(r, \theta)$ is analytic in a neighborhood of the values (r, θ) on g containing (r_0, θ_0). Let us use (r', θ') to represent polar coördinates with B as pole, and let

$$r' = r'(r, \theta), \qquad \theta' = \theta'(r, \theta) \tag{2.4}$$

be the functions of transformation from the coördinates (r, θ) to the coördinates (r', θ') in the region of overlapping of the two coördinate systems. It is easily proved that

$$f(r, \theta) = f'(r', \theta')\cdot\Delta, \tag{2.5}$$

where $f'(r', \theta')$ is the function playing the same rôle for the (r', θ') coördinates as f does for the (r, θ) coördinates, and

$$\Delta = \begin{vmatrix} r'_r & \theta'_r \\ r'_\theta & \theta'_\theta \end{vmatrix}.$$

But by Lemma 1 the functions in (2.4) can have their definition analytically extended even outside the region in which (r, θ) serve as a coördinate system, just so long as (r', θ') serve as a coördinate system. This holds for (r, θ) in a neighborhood of their values on g containing a point C: (r_1, θ_0) just before the first conjugate point to B. These values of (r, θ) yield by (2.4) (extended) values of (r', θ') for which $f'(r', \theta')$ is still analytic. Thus, according to (2.5), $f(r, \theta)$ is analytically continuable for (r, θ) in a neighborhood of their values on g containing a point C just before the first conjugate point to B on g.

Now we introduce polar coördinates (r'', θ'') with C as pole. Under the transformation

$$r'' = r''(r', \theta'), \qquad \theta'' = \theta''(r', \theta'), \tag{2.6}$$

in the region of overlapping of the (r', θ') coördinates and (r'', θ'') coördinates

$$f'(r', \theta') = f''(r'', \theta'') \cdot \Delta'(r', \theta') , \tag{2.7}$$

where

$$\Delta'(r', \theta') = \begin{vmatrix} r''_{r'} & r''_{\theta'} \\ \theta''_{r'} & \theta''_{\theta'} \end{vmatrix} . \tag{2.8}$$

Hence in this region, under (2.4) (extended) and (2.6),

$$f(r, \theta) = f''(r'', \theta'') \cdot \Delta'(r', \theta') \cdot \Delta(r, \theta) . \tag{2.9}$$

Now the resultant of (2.4) (extended) and (2.6) is a transformation

$$r'' = r'' (r, \theta), \qquad \theta'' = \theta'' (r, \theta) . \tag{2.10}$$

Computation shows that

$$\Delta'(r', \theta') \cdot \Delta(r, \theta) = \begin{vmatrix} r''_{r} & r''_{\theta} \\ \theta''_{r} & \theta''_{\theta} \end{vmatrix} = d(r, \theta) . \tag{2.11}$$

Hence, from (2.9)

$$f(r, \theta) = f''(r'', \theta'') \cdot d(r, \theta) \tag{2.12}$$

under (2.10).

But by Lemma 1, (2.10) can be extended to hold for (r, θ) in a neighborhood of their values on g containing a point D: (r_2, θ_0) just before the first conjugate point to C. Values of (r, θ) in this neighborhood yield by (2.10) (extended) values of (r'', θ'') for which $f''(r'', \theta'')$ is still analytic. Thus $f(r, \theta)$ is analytically continuable for (r, θ) in a neighborhood of their values on g reaching almost up to the first conjugate point to C on g. This process can evidently be continued indefinitely. The neighborhood of a pair (r, θ_0) with r arbitrarily large can be reached, because a finite arc of g contains a finite number of points conjugate to A.

But θ_0 was an arbitrary value of θ. Thus $f(r, \theta)$ can be extended to be analytic for all θ and all $r > 0$.

The analyticity of $K(r, \theta)$ is proved more easily due to the invariance of K. Let (r_0, θ_0) be an arbitrary pair with $r_0 > 0$, and let P be the corresponding point of S. Let (x, y) be an admissible coördinate system in the neighborhood of P. Then (x, y) in the neighborhood of their values at P and (r, θ) in the neighborhood of (r_0, θ_0) are connected by an analytic relation

$$x = x(r, \theta) , \qquad y = y(r, \theta) . \tag{2.13}$$

But under (2.13)

$$K(r, \theta) = K'(x, y) ,$$

where $K'(x, y)$ gives the Gaussian curvature of S in terms of (x, y) and is an analytic function of (x, y).

Hence $K(r, \theta)$ is analytic in the neighborhood of the arbitrary pair (r_0, θ_0) and the theorem is proved.

3. **Locus of conjugate points.** Let us now consider a function $f(r, \theta)$ for which the conditions of Theorem 1 are satisfied, i.e., $f(r, \theta)$ is analytic for all θ and all $r > 0$, and so is the function $K(r, \theta) = -f_{rr}/f$. We also assume that the conditions (2.2) are satisfied. We study $f(r, \theta)$ in the euclidean plane of the polar coördinates (r, θ), concerning ourselves with properties which will have significance when we later consider the function f as defining an element E on a complete surface S. Of particular importance are the zeros of f in the (r, θ) plane, since these will give the location of points conjugate to A on S.

Suppose $f(r_0, \theta_0) = 0$, with $r_0 > 0$. We assume that this is the first zero of f after A on the ray $\theta = \theta_0$ issuing from the pole A in the (r, θ) plane.[7] Then $f_r(r_0, \theta_0)$ must be different from zero, since $f(r, \theta_0)$ satisfies (2.3) and is not identically zero.

Hence by the implicit function theorem the equation

$$f(r, \theta) = 0 \tag{3.1}$$

can be solved to give an analytic function

$$r = R(\theta) \tag{3.2}$$

for θ in the neighborhood of θ_0. Since $f_r(r, \theta)$ is $\neq 0$ at all zeros of f, the function (3.2) can be extended, always satisfying (3.1), until either θ reaches a value θ' for which f has no zero except $r = 0$, or θ has increased by 2π and $R(\theta)$ has returned to its value $R(\theta_0)$. In the former case, as $\theta \to \theta'$, $R(\theta) \to \infty$. In the latter case, $R(\theta)$ has a period 2π. In either case, all the pairs (r, θ) defined by (3.2) give *first* zeros of f beyond A on their respective rays from A. In the case where $R(\theta) \to \infty$ as $\theta \to \theta'$, $R(\theta)$ can be extended as θ goes in the other direction until θ reaches a value θ'' for which f has no zero except $r = 0$. θ'' may or may not be identical with θ'. If it is not, then there are in general other intervals for θ in which (3.1) has similar solutions.

We can summarize in the following

LEMMA 2. *Let $f(r, \theta)$ be a function satisfying the conditions* (2.2) *and the conditions of Theorem* 1. *Then the locus in the plane of the polar coördinates (r, θ) of the first zeros of f beyond A on the rays θ = constant issuing from A is either nothing at all, or a single closed curve C about A, or a set of one or more distinct open curves $\{L\}$, each asymptotic to a pair of rays from A.*

In the case of a single closed curve C, unless it is a circle about A as center it will contain a finite number of points N_i nearest to A relative to neighboring points of C, and an equal number of points F_i relatively farthest from A. At each of either of these two types of points $R' = 0$. But $f[R(\theta), \theta] \equiv 0$ and hence

$$f_r R' + f_\theta = 0 \tag{3.3}$$

[7] That the zero of f at $r = 0$ is isolated follows from conditions (2.2).

along C, so that

$$(3.4) \qquad f_\theta = 0$$

at the points in question.[8]

In the case of a set of one or more open curves $\{L\}$, each curve L must contain one or more points N_i. The number of points F_i for each curve L is one less than the number of points N_i for that curve. At each point N_i or F_i $f_\theta = 0$. A curve L can contain an infinite number of points N_i or F_i only if these points recede infinitely far from A. Also the number of curves in the set $\{L\}$ can be infinite only if the set recedes infinitely far from A; for otherwise the points N_i would have a limit point (r_1, θ_1) which would be a zero of f, and yet there would be a ray in every neighborhood of $\theta = \theta_1$ which would have no zero of f but $r = 0$. This gives a contradiction.

Thus we have another lemma.

LEMMA 3. *The curve C of Lemma 2 contains a finite number of points N_i relatively nearest to A and an equal number of points F_i relatively farthest from A, unless it is a circle with A as center. In the case of a set of curves $\{L\}$, each curve of the set contains at least one point N_i and a number of points F_i one less than the number of points N_i. The number of these points on a single curve of the set $\{L\}$ can be infinite only if the points recede infinitely far from A. At all these points $f_\theta = 0$. The number of curves in the set $\{L\}$ can be infinite only if the set recedes infinitely far from A.*

Now suppose that we have an element E about A on a complete surface S. Then by Theorem 1, the hypotheses of Lemmas 2 and 3 are satisfied. The pairs (r, θ) which give the first zeros beyond A of the function f taken along the rays issuing from A in the (r, θ) plane locate for us the first points conjugate to A on the geodesic rays issuing from A on S.

The locus of these points on S conjugate to A will be given locally by equations of the form

$$(3.5) \qquad x = x[R(\theta), \theta] = X(\theta), \qquad y = y[R(\theta), \theta] = Y(\theta),$$

where x and y form a local admissible coördinate system on S and $x(r, \theta)$, $y(r, \theta)$ are analytic functions of (r, θ). Then X and Y are analytic functions of θ.

From Lemma 2 we obtain

THEOREM 2. *On a complete surface S the locus of the first points conjugate to A on the geodesic rays issuing from A is either no point at all, or a closed curve C (perhaps a single point), or a set of one or more open curves $\{L\}$. In each case the locus can be analytically parametrized in terms of θ.*

Let us study the functions (3.2). Denoting the Riemannian metric in the coördinates (x, y) by $Adx^2 + 2Bdxdy + Cdy^2$ we have

$$(3.6) \qquad Adx^2 + 2Bdxdy + Cdy^2 = dr^2 + f^2d\theta^2$$

[8] It can also be proved that at the points N_i $f_{\theta\theta} \geqq 0$ and at the points F_i $f_{\theta\theta} \leqq 0$. Similarly, the higher tests for maxima and minima give analogous formal results.

under the transformation

$$x = x(r, \theta), \qquad y = y(r, \theta). \tag{3.7}$$

We find that

$$Ax_r^2 + 2Bx_ry_r + Cy_r^2 = 1, \qquad Ax_\theta^2 + 2Bx_\theta y_\theta + Cy_\theta^2 = f^2. \tag{3.8}$$

Hence x_r, y_r are not both zero, and at a point conjugate to A lying in the region of S in which (x, y) form an admissible coördinate system

$$x_\theta = 0, \qquad y_\theta = 0. \tag{3.9}$$

Upon differentiating equations (3.5) we find that

$$X' = x_rR' + x_\theta = x_rR', \qquad Y' = y_rR' + y_\theta = y_rR'. \tag{3.10}$$

Hence along the locus of conjugate points

$$\frac{dy}{dx} = \frac{y_r}{x_r}, \tag{3.11}$$

and the locus has a continuously turning tangent (since x_r and y_r are nowhere both zero). Furthermore, it follows from (3.10) that the locus is the envelope of the geodesic rays from A.

If $R' \equiv 0$, (3.10) tells us that $X' \equiv Y' \equiv 0$. This means that the locus is a single point. Conversely, if the locus is a single point, $R' \equiv 0$.

Now suppose $R' \not\equiv 0$. Then the irregular points of the locus or its representation (3.5) (the points at which $X' = Y' = 0$) are given by the zeros of R'. Among these points are the points N_i and F_i of Lemma 2. At these points R has minimum and maximum values, respectively, and hence R' changes sign. From (3.10) we then deduce that the locus has a cusp at each of these points.

Using these results and Lemma 3, we obtain

THEOREM 3. *The curve C of Theorem 2 (if it is not a single point) contains a finite number of cusps turned toward A and an equal number turned away from A.*[9] *In the case of a set of open curves $\{L\}$, each curve of the set is infinitely long, and contains at least one cusp turned toward A, and one fewer cusps turned away from A. A single curve of $\{L\}$ cannot contain an infinite number of cusps on a finite segment of the curve. The number of curves in the set $\{L\}$ can be infinite only if a curve of the set can be found all of whose points are arbitrarily far from A along the geodesics on which they are conjugate to A.*

4. **Locus of minimum points.** The importance of minimum points, as defined in the introduction, is suggested by the following theorem proved by Rinow:

(D) Through every point on a complete surface homeomorphic to the plane

[9] Carathéodory has proved that for an "Eifläche", the curve C has at least two cusps of each kind. See Blaschke, loc. cit., p. 103. Carathéodory's proof holds for any Riemannian surface which has a closed curve as the locus of first conjugate points to A.

there is at least one geodesic ray every segment of which furnishes an absolute minimum to the arc length of curves joining its end points.

Such a geodesic ray (or geodesic arc) will be said to be *of class* $\mathfrak{A}$.

On a surface S homeomorphic to the sphere there can be no geodesic ray of class $\mathfrak{A}$. For there can be no set of points on S without a limit point on S, and on a geodesic ray of class $\mathfrak{A}$ there is such a set of points. Hence we have the following

THEOREM 4. *A complete simply connected surface S is homeomorphic to the plane if and only if it contains a geodesic ray of class $\mathfrak{A}$ issuing from an arbitrary point A. It is homeomorphic to the sphere if and only if it contains no geodesic ray of class $\mathfrak{A}$.*

If a geodesic ray issuing from A is not of class $\mathfrak{A}$, it is easily seen that there must be a minimum point on g with respect to A, as defined in the introduction to this paper. The importance of a study of the locus of minimum points with respect to A on the geodesic rays issuing from A is shown by Theorem 4, which gives us an essential difference between the locus on a surface homeomorphic to the sphere and on a complete surface homeomorphic to the plane. In a later paper it will be shown that the topology of any complete Riemannian surface is completely determined by a knowledge of the properties of the minimum point locus.

We now give several useful lemmas.[10]

LEMMA 4. *Two geodesic arcs of class $\mathfrak{A}$ cannot cross more than once; they cannot even meet once and cross once.*

For if they did, there would exist a geodesic arc of class $\mathfrak{A}$ with a corner.

Using the same notation as in Lemma 1 used in the proof of Theorem 1 of §1, we have the following:

LEMMA 5. *If P_i: $(r_i, \theta_i) \to P$ on S, and (r_0, θ_0) is a limit pair of values of (r_i, θ_i), (r_0, θ_0) represents P.*

This follows from Lemma 1.

A geodesic arc g through A is said to be a *limit geodesic arc* of a set of geodesic arcs through A if there exists a sequence g_i of the set such that the initial directions of g_i and the lengths of g_i approach as respective limits the initial direction and length of g. Lemma 1 tells us that the terminal points of g_i approach as a limit the terminal point of g. We can now prove

LEMMA 6. *A limit geodesic arc g through A of a set of geodesic arcs of class $\mathfrak{A}$ through A is also of class $\mathfrak{A}$.*

Suppose that g of length l were not of class $\mathfrak{A}$; let $\bar{g}$ of length $\bar{l}$ be a shorter arc than g joining the end points A and B of g. Suppose

$$l - \bar{l} = 2\epsilon. \tag{4.1}$$

Let g_i be the sequence of geodesic arcs of class $\mathfrak{A}$ approaching g as a limit. For k sufficiently large, the geodesic arc g_k has its end point B_k within a distance ϵ of

[10] These lemmas are not restricted to simply connected surfaces.

B, and its length l_k differing from l by less than ϵ. Denote by b a geodesic arc of class $\mathfrak{A}$ from B to B_k. We also use b to denote the length of this arc. Then

$$b < \epsilon, \qquad |\, l - l_k \,| < \epsilon. \tag{4.2}$$

Combining (4.1) and (4.2) we obtain

$$\bar{l} + b < l_k. \tag{4.3}$$

Thus the length of the broken geodesic $\bar{g}b$ is less than that of g_k, which contradicts the fact that g_k is of class $\mathfrak{A}$. Thus the lemma is proved.

LEMMA 7. *A limit geodesic arc g through A of a set of geodesic arcs through A not of class $\mathfrak{A}$ either is not of class $\mathfrak{A}$ or its final end point is a minimum point on it with respect to A.*

Some sequence g_i of the set approaches g as a limit. Let $\bar{g}_i$ be a geodesic of class $\mathfrak{A}$ joining the end points of g_i. The sequence $\bar{g}_i$ must have at least one limit geodesic arc, and this arc has as its end points the end points of g, and is of class $\mathfrak{A}$ by the previous lemma. If g is not itself such a limit geodesic arc, then the lemma is easily seen to be true. If g is a limit geodesic arc of the sequence $\bar{g}_i$, let $\bar{h}_i$ be a subsequence of $\bar{g}_i$ which approaches g, and let h_i be the subsequence of g_i joining the end points of $\bar{h}_i$. Then since $h_i \to g$ and $\bar{h}_i \to g$, if we take an arbitrarily small neighborhood n_1 of the final end point of g, an arbitrarily small neighborhood n_2 of the initial direction of g, and an arbitrarily small neighborhood n_3 of the value of the arc length of g, we can always find a point in n_1 which can be joined to A by two distinct geodesic arcs both with initial directions in n_2 and lengths in n_3. This means that the final end point of g must be conjugate to A and hence is a minimum point on g with respect to A. This completes the proof of the lemma.

LEMMA 8. *If a sequence of points $\bar{M}_i$ which are minimum points with respect to A on geodesics $\bar{g}_i$ issuing from A approaches as a limit a point M, then M is a minimum point with respect to A on every limit geodesic of the sequence $\bar{g}_i$.*

By Lemma 5, any limit geodesic arc $\bar{g}$ of $\bar{g}_i$ will pass through M. By Lemma 6, the arc AM of $\bar{g}$ is of class $\mathfrak{A}$. Now let P be a point on $\bar{g}$ beyond M but arbitrarily close to it. The arc AP on $\bar{g}$ is a limit geodesic arc of a set of geodesic arcs not of class $\mathfrak{A}$, and by Lemma 7 either AP is not of class $\mathfrak{A}$ or P is a minimum point on $\bar{g}$ with respect to A. Thus the minimum point on $\bar{g}$ with respect to A is at or before P. But since P is arbitrarily close to M, M itself must be a minimum point on $\bar{g}$ with respect to A, and the lemma is proved.

LEMMA 9. *If M is a minimum point on g with respect to A, and g is the only geodesic arc of class $\mathfrak{A}$ joining A to M, then M is conjugate to A on g.*

Consider a set of points P_i on g extended, beyond M and approaching M. Each point P_i can be joined to A by a geodesic arc g_i of class $\mathfrak{A}$ shorter than AP_i on g. Then by Lemma 6 and our hypothesis, g is the limit geodesic arc of g_i. Thus in every neighborhood of M the geodesic arc g intersects another geodesic ray from A arbitrarily close to g (in initial direction and length). This proves the lemma.

LEMMA 10. *If M is a minimum point on g with respect to A, and M is conjugate to A on g, then M is a cusp turned toward A of the locus of first points conjugate to A, or M itself is the complete locus of points conjugate to A.*

For if M is not the complete locus of points conjugate to A, that locus is described by Theorems 2 and 3. M must be a cusp turned toward A of that locus; otherwise by the Envelope Theorem[11] the arc AM on g would be replaceable by a shorter arc joining A to M.

In general, a minimum point M with respect to A can be joined to A by two or more geodesics of class $\mathfrak{A}$ of equal length. The number of geodesic arcs of class $\mathfrak{A}$ joining A to a minimum point M with respect to A is called the *order* of the minimum point.

LEMMA 11. *If g and $\bar{g}$ are two geodesic rays issuing from a point A on S and meeting for the first time at a point M so that the arc AM has the same length on g and $\bar{g}$, and so that M is not conjugate to A on either g or $\bar{g}$, then there is a regular analytic arc d through M such that each point of d is the intersection of two geodesic rays of equal length issuing from A, one in the neighborhood of g and the other in the neighborhood of $\bar{g}$. Furthermore, every point with this property lies on d.*

Since M is not conjugate to A on g or $\bar{g}$, the geodesic rays through A neighboring g simply cover the neighborhood of M, and so do those neighboring $\bar{g}$. Thus we have two admissible coördinate systems (r, θ) and (r', θ') in the neighborhood of M, and they are connected by a non-singular analytic correspondence

$$r' = r'(r, \theta), \qquad \theta' = \theta'(r, \theta). \tag{4.4}$$

Now at the point M clearly $r'_r \neq 1$, since g and $\bar{g}$ meet at a non-zero angle. Hence we can solve the equation

$$r'(r, \theta) - r = 0 \tag{4.5}$$

for r as an analytic function of θ, $r = r(\theta)$, in the neighborhood of $\theta = \theta_0$, the θ-coördinate of g. Using the functions $x(r, \theta)$ of Lemma 1 in the neighborhood of M, we obtain an analytic arc

$$x = x[r(\theta), \theta] = x(\theta), \qquad y = y[r(\theta), \theta] = y(\theta).$$

This arc is regular, for the derivatives $x'(\theta)$ and $y'(\theta)$ are both zero only if

$$\begin{vmatrix} x_r & x_\theta \\ y_r & y_\theta \end{vmatrix} = 0,$$

which is contrary to the hypothesis that M is not conjugate to A.

We can now go about giving a complete description of the locus m of minimum

[11] The Envelope Theorem states that the change in arc length along the conjugate point locus is equal to the change in r along the locus. It follows immediately from the fact that along the conjugate point locus $ds^2 = dr^2$.

points with respect to A. *We assume henceforth that the surface S is simply connected.*

First we consider the case that S is closed, i.e., homeomorphic to the sphere. To every value of θ corresponds one minimum point, and using Lemma 8 we conclude that the locus m is the continuous single-valued image of a circle.

Let us dispose of the possibility that m consists entirely of a single point M. Such a point is obviously conjugate to A on every geodesic ray through A, and is equidistant from A on every such geodesic ray. Thus this case coincides with the case already mentioned in Theorems 2 and 3, when the locus C of first conjugate points to A is a single point. Even if we know only that all the geodesic rays through A for some small interval $\theta_1\theta_2$ have their minimum points with respect to A at M, we can conclude from the analyticity of the conjugate point locus that m consists only of M. We rule out this case in the following work.[12]

On a closed S, m cannot contain any simple closed curve C. For C would divide S into two regions R_1 and R_2 such that no point of R_1 could be joined to a point in R_2 without crossing C. One of these, say R_1, would contain A. Since S is complete, A could be joined to any point in R_2 by a geodesic of class $\mathfrak{A}$. But this geodesic would have to cross C. This contradicts the fact that it is of class $\mathfrak{A}$.

By an *end point* of m we shall mean a point P of m from which there issues one and only one 1-cell composed of points of m. The object of the next two paragraphs is to prove that m contains only a finite number of end points. This, together with the absence of closed curves in m, will prove that m is a tree (i.e., a finite 1-dimensional complex without closed curve).[13]

In the first place, every end point P of m is a minimum point of order 1. For if P could be joined to A by $n > 1$ geodesics of class $\mathfrak{A}$, $n - 1$ simple closed curves would be formed dividing S into n bounded 2-cells $S_1, \cdots, S_n$. Let g: $\theta = \theta_0$ and g': $\theta = \theta_0'$ be the geodesic rays bounding the 2-cell S_1. Then every ray $\theta = \theta_1$ for $\theta_0 < \theta_1 < \theta_0'$ contains in S_1 its minimum point with respect to A. For if this ray crosses g or g', it must cease to be of class $\mathfrak{A}$, and if it does not cross g or g', but remains in the interior of S_1 for its whole infinite length, the boundedness of S_1 shows that it cannot be of class $\mathfrak{A}$ along its whole infinite length. A similar situation holds in $S_2, \cdots, S_n$. Thus at least n continuous curves of m issue from P, one in each S_i. Hence P is not an end point of m if the order of P as a minimum point is > 1, and so every end point of m is a minimum point of order 1.

There are only a finite number of minimum points of order 1 with respect to A on S. For by Lemma 9 a minimum point of order 1 will be conjugate to A, and by Lemma 10 it will be a cusp of the locus of first conjugate points to A.

[12] Examples of this case occur when A is any point on a sphere or one of the poles of an ellipsoid of revolution.

[13] This conclusion is a well-known fact in curve theory. For example, it can easily be proved from either of two theorems given by K. Menger in his book *Kurventheorie*, one on p. 266, and the other on p. 323.

By Theorem 3, there cannot be an infinite number of cusps of the locus of first conjugate points to A unless the cusps recede infinitely far from A along the geodesics on which they are conjugate to A, and this is impossible on the bounded surface S for cusps which are minimum points with respect to A on the geodesics on which they are conjugate to A.

Hence there are only a finite number of end points of m, and m is a tree. We now show that the order of a point of m as a minimum point with respect to A is always equal to its order as a vertex of m.[14]

To identify the end points of m with the minimum points of order 1, it remains to prove that every minimum point of order 1 is an end point of m. This we do in the next three paragraphs.

Suppose that two geodesics of class $\mathfrak{C}$ g: $\theta = \theta_0$ and g': $\theta = \theta_0'$ join A to a point M. As proved previously, the bounded 2-cell S_1 formed by these two arcs and containing the rays $\theta_0 < \theta < \theta_0'$ contains all the minimum points on these rays. All the geodesic rays on an interval $\theta_1\theta_2$ in the interval $\theta_0\theta_0'$ cannot have their minimum points with respect to A at M, for that is the exceptional case already ruled out. Let us then draw a geodesic ray h near g from A into S_1 whose minimum point with respect to A is not M. If this minimum point is not of order 1, h will meet an equally long ray h' from A at the minimum point. The arcs h and h' form a simple closed curve lying in S_1. Continuing this process, we see that at least once in S_1 these pairs of arcs reach the limiting case of a single geodesic arc k from A. The minimum point on k with respect to A will be a minimum point of order 1. Thus S_1 contains at least one minimum point of order 1, and it is on a geodesic ray whose θ lies in the interval $\theta_0\theta_0'$.

If S_1 contains just one minimum point N of order 1 with respect to A, no point in S_1 is a minimum point of order > 2. For otherwise a repetition of the process of the previous paragraph would furnish at least two minimum points of order 1 in S_1. The value of θ, say θ', for which N occurs divides the interval $\theta_0\theta_0'$ into two intervals whose points (i.e., values of θ) are in a one-to-one continuous correspondence in which each pair of correspondents is a pair of values of θ furnishing a pair of geodesic rays whose intersection is a minimum point on both rays with respect to A. Hence the locus in S_1 of minimum points with respect to A is, by Lemma 8, given by either of two continuous functions

$$r = r_1(\theta) \qquad (\theta_0 \leqq \theta \leqq \theta'),$$

$$r = r_2(\theta) \qquad (\theta' \leqq \theta \leqq \theta_0').$$

By Lemma 1, the locus can be represented locally by equations of the form

$$x = x[r_1(\theta), \theta], \qquad y = y[r_1(\theta), \theta],$$

where these functions are continuous. The locus cannot cross itself, for a multiple point would be a minimum point of order > 2. Hence the locus in S_1 is a

[14] By the *order of P as a vertex of m* we mean the number of 1-cells of m incident with P.

1-cell with M and N as end points. The minimum point N of order 1 is thus an end point of m.

Now let P' be an arbitrary minimum point of order 1 with respect to A, on the geodesic ray $\theta = \theta'$. Since the number of minimum points of order 1 is finite, there exists an interval I: $\theta_1 < \theta < \theta_2$ containing θ' but no other value of θ which furnishes a minimum point of order 1. Each value of θ in this interval furnishes a minimum point of order >1, and hence to each such value of θ corresponds another value of θ, say θ'', which furnishes the same minimum point P. If θ is close enough to θ', then P is very close to P', by Lemma 8, and from Lemma 2 and the fact that P' is of order 1 we conclude that θ'' is very close to θ'. Thus if θ is close enough to θ', both θ and θ'' lie in $\theta_1\theta_2$. Since the interval $\theta\theta''$ must include a value of θ furnishing a minimum point of order 1 (see paragraph before previous one), we have $\theta_1 < \theta < \theta' < \theta'' < \theta_2$. Thus the rays θ and θ'' form a 2-cell S_1 of the kind discussed in the previous paragraph containing just one minimum point of order 1, and hence P' is an end point of m.

But P' was an arbitrary minimum point of order 1, so that every minimum point of order 1 is an end point of m. We have already proved that every end point of m is a minimum point of order 1, so that these two sets of points are identical.

We now prove by induction that a vertex of order n of the tree m is a minimum point of order n and conversely. We have already proved this for $n = 1$. Assume the proposition true for all integers $n < q$; we shall prove it for $n = q$. A vertex P of m of order q cannot be a minimum point of order $w > q$; for we have already seen that from the latter type of point issue at least w distinct 1-cells of the tree m. Furthermore, P cannot be a minimum point of order $< q$; for part of our induction hypothesis is that a minimum point of order $< q$ is a vertex of m of the same order. Thus the order of P as a vertex of the tree m equals its order as a minimum point with respect to A.

This completes the case where S is closed. If S is open, that is, homeomorphic to the plane, at least one geodesic ray from A contains no minimum point with respect to A. The locus m can, by methods similar to those already used, be shown to consist of one or more distinct trees each of which contains a branch which extends infinitely far from A. Each tree may have an infinite number of branches issuing from this branch, but not on any bounded segment of it. As in the case of a closed S, the order of a point M of m as a vertex of m equals its order as a minimum point with respect to A; also the end points of m are conjugate to A, and are cusps of the locus of first conjugate points to A.

Whether S is open or closed, if we cut off each geodesic ray through A at its minimum point, the region $S - m - A$ is simply covered by these truncated geodesic rays. For two such geodesic rays cannot intersect before m; if they did, the absolute minimum would have stopped on at least one of them at or before the intersection with the other, which gives a contradiction.

Finally, Lemma 11 tells us that any arc of m containing no points conjugate to A and no interior points of order > 2 is a regular analytic arc.

We summarize in the following

THEOREM 5. *Let A be an arbitrary point on the complete simply connected surface S. If S is homeomorphic to the sphere, the locus m of minimum points with respect to A consists of a single point, or a tree with a finite number of branches. If m is a single point, it is conjugate to A on all the geodesics joining it to A, and at equal distances from A along these geodesics. If S is homeomorphic to the plane, m consists of nothing at all, or a set of one or more distinct trees each of which contains one branch extending infinitely far from A. There may be an infinite number of branches extending from this branch, but not from any bounded segment of it. Whether S is open or closed, the order of a point M of m as a vertex of m is equal to the order of M as a minimum point with respect to A. The end points of m are conjugate to A, and are cusps of the locus of first conjugate points to A. An arc of m containing no points conjugate to A and no interior points of order > 2 is a regular analytic arc. The region $S - m - A$ is simply covered by the geodesic rays through A cut off at their intersections with m, and hence $S - m$ forms a two-cell with m as its singular boundary, in which the geodesic polar coördinates with A as pole form a coördinate system.*

This reduction of a complete surface to a single 2-cell with a linear graph as its singular boundary is analogous to the process in analysis situs of making a similar reduction by coalescing 2-cells of a 2-dimensional manifold. See, for example, Veblen's *Analysis Situs*, p. 74 §§62, 63. What we have in this paper is a method of associating with each point of a simply connected surface a specific tree, or set of trees, which bounds the 2-cell formed by the geodesic rays issuing from A.

5. **Examples.** We have already mentioned examples in which the minimum point locus is a single point. These are the sphere, in which the locus with respect to any point A is a single point, and the ellipsoid of revolution, in which the locus with respect to either pole is the opposite pole. On the general ellipsoid, the locus with respect to any of the six poles is an arc on the longer of the two principal ellipses through the given pole, containing the opposite pole as midpoint. In the euclidean or hyperbolic plane, the locus with respect to any point is non-existent; the same result is true on any simply connected developable surface. Examples of well-known simply connected surfaces on which the minimum point locus assumes a complicated, but determinable, form are naturally hard to give.

6. **Generalizations and problems.** The author has already generalized this work to closed multiply connected surfaces,[15] the fundamental theorem being stated in the introduction of this paper. The details will appear in a later paper, together with a treatment of the non-analytic case. Another question of interest is a study of the variation of the minimum point locus as the point A moves over the surface S. In this connection, we have the well-known problem

[15] See Myers, Proceedings of the National Academy of Sciences, April, 1935, Theorem 5.

of Blaschke,[16] which may now be stated as the problem of finding all the surfaces such that the locus of minimum points with respect to every point A is a single point.

It is likely that the minimum point locus appears in n-dimensional Riemannian manifolds. Also, since the minimum point locus can be expressed as the locus of a certain type of intersection of the paths issuing from A, it may be that a similar locus exists in spaces of the geometry of paths.

Added in proof (Aug. 22). J. H. C. Whitehead has been kind enough to call my attention to a paper by Poincaré of which I had previously been unaware. See Trans. Am. Math. Soc. (1905), pp. 237–274. In this paper Poincaré devotes p. 243 to a discussion of what I have called the locus of "minimum points" and which he calls "lignes de partage", restricting himself to closed simply connected analytic surfaces of positive curvature. He states that the locus cannot separate the surface and hence cannot contain a closed curve, that the end points of the locus are conjugate to the original point A and are cusps turned toward A of the locus of first conjugate points to A, and if n shortest geodesics join a point M of the locus to A, then n 1-cells of the locus issue from M and bisect the angles between consecutive geodesics. Also, a paper by J. H. C. Whitehead has just appeared in the July number of the Annals of Mathematics which is closely connected with the present one.

PRINCETON UNIVERSITY AND THE INSTITUTE FOR ADVANCED STUDY.

[16] See Blaschke, loc. cit., pp. 224–230, §102.

ON CRITICAL SETS

By S. Lefschetz

A number of topological investigations of recent years have had for objective the freeing of algebraic topology from extraneous considerations and hypotheses. The most noteworthy are perhaps the endeavors to obtain an intrinsic theory of manifolds. The starting point has been to replace the regularity typified by euclidean spaces by conditions expressed exclusively in terms of chains and homologies. Our present purpose is to exploit our recent results on chain-deformations[1] to bring Morse's important theory of critical points[2] within the framework of algebraic topology. The treatment will be found to be substantially free from the difficulties usually connected with homotopy.

That the free use of topology more than justifies itself will immediately be perceived from the examples given in No. 2 and from the treatment of critical points of functions on manifolds which occupies the last section of the paper.

§1. Critical values of a function on a set

1. Let $f(x)$ be a real continuous function of the point x on a metric space $\mathfrak{R}$. We propose to investigate the variation of the homology structure of the spaces $a \leqq f < y$, $a \leqq f \leqq y$ as y varies. The range taken at any one time shall be from some finite a up, although the restriction is relatively unimportant. In fact, the range might well be the whole real axis or the real numbers mod 1 (circumference) without entailing essential modifications in the treatment.

Upon examination, Morse's theory is found to demand essentially (a) chain-deformation downwards across all but some isolated sets $y =$ constant, and also downwards away from all these sets without exception; (b) the finiteness of the homology characters (type numbers). Our structural axioms practically amount to imposing these properties.

Received June 14, 1935. The results of the present paper were communicated in a note in the Proceedings of the National Academy, vol. 21 (1935), pp. 220–222.

[1] See S. Lefschetz, *Chain-deformations in topology*, Duke Journal, vol. 1 (1935), pp. 1–19 (= DJ in the sequel). The assumptions, notations and mode of reference of DJ will be used in the present paper. In particular, *Topology* designates our Colloquium Lectures. We also recall that the symbols HLC, HNR stand, respectively, for "locally connected" and "neighborhood retract" in the sense of homology.

[2] The chief references (prior to 1934) to the very extensive literature, chiefly due to Morse and his students, which has grown up in recent years around these questions will be found at the end of Morse's Colloquium Lectures, *The Calculus of Variations in the Large*, New York, 1934, referred to as MC in the sequel. In addition, we may mention our recent note in the Proceedings of the National Academy, and a subsequent note by Morse-Van Schaack, ibid., pp. 258–263.

It will be expedient to use the following notation

$$U^y = \text{the open set } f(x) < y;$$

$$A^y = \overline{U^y} = \text{the set } f(x) \leqq y;$$

$$B^y = \text{their common boundary, the set } f(x) = y.$$

If we think of y as a "vertical" coördinate, B^y is the section of $\mathfrak{R}$ by a "hyperplane" y = constant, U^y the part of $\mathfrak{R}$ below the hyperplane, A^y the part of $\mathfrak{R}$ not above the hyperplane.

The space axioms. AXIOM I. *Every A^y is compact and* HLC.

AXIOM II. *Every U^y is shrinkable away from any point of B^y.*

We propose to investigate the variation of the homology groups and characters of the sets A^y and U^y when our axioms hold. Before doing so we shall first present a few examples.

2. **Examples.** I. $\mathfrak{R}$ is a bounded p-dimensional polyhedron in a euclidean n-space S_n, and $f(x) = y$ is one of the coördinates. The sets A^y are all finite dimensional polyhedra; hence Axiom I is fulfilled by virtue of Theorem XII of DJ. As for Axiom II, we may subdivide A^y into the cells of a simplicial polyhedral complex K in such a manner that B^y coincides with a subcomplex L of K. Since we may replace K by any subdivision, we may assume the complex such that the K-neighborhood N, of L, is *normal* in the sense of *Topology*, p. 91. The characteristic property of such a neighborhood is the following (see loc. cit.). If Φ is the boundary of N, through every point P of $N - L$ there passes a unique segment QR with $Q \subset \Phi$, $R \subset L$. The homotopy $P \to Q$ along the segment PQ induces a chain-deformation of $\bar{N} - L$ onto $K - N$ (DJ, Th. I), i.e., a chain-shrinking of U^y away from $L = B^y$ over $K = A^y$, so that Axiom II holds.

II. $\mathfrak{R}$ is a bounded analytical locus in S_n, f an analytic function regular at all points of $\mathfrak{R}$. Here every A^y is likewise an analytical locus, and hence it may be covered with a simplicial complex.[3] We are therefore back to the preceding case.

III. $\mathfrak{R}$ is a compact analytical locus in one of the following spaces: (a) real projective space; (b) complex projective space; (c) a topological analytical $\mathcal{M}_n$, i.e., a locally euclidean n-space with allowable coördinate systems mutually analytical wherever they overlap. The function f is an analytic point function on the space or on $\mathcal{M}_n$, regular over $\mathfrak{R}$. For the covering by complexes may be extended to these three cases as in *Topology*, p. 378.

A noteworthy special case: $\mathfrak{R}$ is a real or complex algebraic variety.

IV. $\mathfrak{R}$ is the closure of a region in one of the spaces just considered whose

[3] See, for example, Lefschetz-Whitehead, *On analytical complexes*, Trans. Amer. Math. Soc., vol. 35 (1933), pp. 510–517.

boundary is a topological manifold of class C''; f is a function of class C''[4] over $\mathfrak{R}$ such that the subsets of $\mathfrak{R}$ for which $f(x) \leqq$ constant are all bounded. We have then the problem of the critical points of functions investigated at length by Morse. We shall consider it more fully in §5.

In the first three examples we could as well have substituted class C'' for analyticity wherever it occurs.

3. We have immediately from Axiom I and DJ, Th. VIII,

THEOREM I. *The homology groups of A^y have the same structure as for a finite complex.*

For our further progress we shall also find very convenient the

LEMMA. *If $\mathfrak{R} \supset A \supset B$, B compact, and if A may be chain-shrunk onto B over itself, then the homology groups of A and B are the same.*[5]

Let $\Gamma = \{\gamma_p\}$ be a homology class of A. By assumption, there is a chain-deformation ϑ of γ_p onto B, and if $\mathfrak{T}\gamma_p$ is the deformation-chain

$$\mathfrak{T}\gamma_p \rightarrow \gamma_p' - \gamma_p; \qquad \gamma_p' \sim \gamma_p \qquad \text{on } A.$$

Hence $\Gamma_p \supset \gamma_p'$. If γ_p^* is another cycle of Γ_p and it is deformed into $\gamma_p^{*\prime}$, we have

$$\gamma_p^{*\prime} \sim \gamma_p^* \sim \gamma_p \sim \gamma_p' \qquad \text{on } A;$$
$$A \supset c_{p+1} \rightarrow \gamma_p' - \gamma_p^{*\prime} \subset B.$$

Since B is compact, we may chain-deform c_{p+1} onto B without changing its boundary (DJ, Th. II); hence

$$B \supset c_{p+1}' \rightarrow \gamma_p' - \gamma_p^{*\prime} \sim 0 \qquad \text{on } B.$$

Therefore Γ_p determines a unique homology class Γ_p' on B. In particular, if $\gamma_p \sim 0$ on A, both γ_p^* and $\gamma_p^{*\prime} = 0$, and hence $\gamma_p' \sim 0$ on B. Hence $\Gamma = 0$ determines $\Gamma' = 0$ on B.

Conversely, if $\gamma_p', \gamma_p^{*\prime}$ are in the same homology class Γ' as to B, they are homologous on A and hence in the same homology class on A. Therefore the correspondence $\Gamma \leftrightarrow \Gamma'$ is one-to-one and zero-preserving. It is also manifestly linear; hence it is a simple isomorphism between the p^{th} homology groups. This proves the lemma.

[4] A function is said to be of class C^r whenever it is continuous together with all its partial derivatives of order $\leqq r$. A topological manifold $\mathcal{M}_n$ is of class C^r whenever the transformations between admissible coördinate systems are given by functions of class C^r wherever their ranges overlap. These notions will be found more fully developed in Veblen-Whitehead, *The Foundations of Differential Geometry*, Cambridge Tract, 1932.

[5] The proof is essentially the same as given by A. B. Brown for a more restricted situation in his paper *Critical sets of an arbitrary analytic function of n variables*, Annals of Math., vol. 32 (1931), p. 514.

4. We pass now to the direct study of the homology groups of the sets A, U.

Under Axiom I every A^y is the subset of some compact $A^{y'}$, $y' > y$, or it is $\mathfrak{R}$, which is then compact. Hence owing to Axiom I and Th. IX of DJ, A^y is an HNR. Since $f(x)$ is continuous on some compact set $\supset A^y$, the HNR property of A^y implies that for every $\epsilon > 0$ there is an $\eta(\epsilon, y)$ such that $A^{y+\eta}$ is ϵ chain-deformable over $U^{y+\epsilon}$ onto A^y. Let us set $\eta_1 = \eta(\eta, y)$, $\eta_2 = \eta(\eta_1, y)$, and let $c_p \subset U^{y+\eta_2}$. The chain is deformable over $U^{y+\eta_1}$ into a $c_p' \subset A^y$, or

$$(4.1) \qquad U^{y+\eta_1} \supset d_{p+1} \rightarrow c_p' - c_p - d_p; \qquad d_p \text{ and } d_{p+1} \subset U^{y+\eta_1}.$$

Since $| c_p |$ is closed, some $A^{y'} \supset | c_p |$, where $y < y' < y + \eta_2$, and the same $A^{y'} \supset | c_p' |, | F(c_p) |, | F(c_p') |$.

Now d_p is chain-deformable over $U^{y+\eta}$ into a chain of A^y, and since $A^{y'} \subset U^{y+\eta_1}$, d_p is chain-deformable over $U^{y+\eta}$ into a $d_p' \subset A^{y'} \subset U^{y+\eta_2}$ and with the same boundary. Since $F(d_p) = F(c_p') - F(c_p)$, we have

$$U^{y+\eta} \supset d_{p+1}' \rightarrow d_p' - d_p,$$

$$(4.2) \qquad U^{y+\eta} \supset d_{p+1} + d_{p+1}' = d_{p+1}'' \rightarrow c_p' - c_p - d_p' \subset U^{y+\eta_1}.$$

If we compare (4.2) with (4.1), we see that it is of the same form as far as c_p, c_p' are concerned, save that now $d_p \subset U^{y+\eta_2}$, and $d_{p+1} \subset U^{y+\eta}$. Starting then again with (4.1) and this new situation, we observe that $F(d_{p+1}) \subset A^{y'} \subset U^{y+\eta_2}$. We may now chain-deform d_{p+1} over $U^{y+\epsilon}$ into a chain $d_{p+1}' \subset A^{y'} \subset U^{y+\eta_2}$ without touching its boundary. Therefore in (4.1) we may actually assume both d_p and $d_{p+1} \subset U^{y+\eta_2}$. This shows that c_p is chain-deformable into c_p' of A^y over $U^{y+\eta_2}$. In other words, $U^{y+\eta_2}$ may be chain-shrunk onto A^y. It follows that any $A^{y'}$ or $U^{y'}, y < y' < y + \eta_2$ may be chain-shrunk onto A^y. From this and the lemma we deduce

(a) When y' is on the open interval $(y, y + \eta_2)$, the homology groups of $A^{y'}$ and $U^{y'}$ are fixed and they are the same as those of A^y, and hence they have the same structure as for a finite complex (Th. I).

(b) Under the same circumstances $A^{y'}$ may be chain-shrunk away from $B^{y'}$.

Property (b) is a consequence of the fact that if c_p of our discussion is in $A^{y'}$, we may choose that $A^{y'}$ as the $A^{y'}$ of the discussion. Since the ultimate chains $d_{p+1}, d_p \subset A^{y'}$, we see that c_p can be chain-shrunk over $A^{y'}$ onto $A^y \subset A^{y'} - B^{y'}$.

5. So far we have only utilized Axiom I. Passing to Axiom II, we observe that since B^y is closed and A^y compact, U^y may be chain-shrunk away from B^y (DJ, Th. III). Therefore U^y is chain-deformable over A^y to the exterior of some open set (relatively to A^y) $\supset B^y$, and hence to the exterior of some $\mathfrak{R} - A^{y'}$ on A^y, or finally onto some $A^{y'}$, $y' < y$. Hence by the lemma, with $y - y' = \zeta$, observing that U^y and $A^{y''}$, $U^{y''}$, $y' < y'' < y$, are chain-shrinkable onto $A^{y'}$, and hence away from $B^{y''}$ on themselves, we conclude

(c) Properties (a), (b) hold also on the open interval $(y - \zeta, y)$, except that the fixed homology groups are those of U^y.

To sum up, we have

THEOREM II. *Every value y is interior to an interval such that* (a) *on both sides of y the sets A, U have fixed homology groups, the same for both, which are those of A^y above, those of U^y below*; (b) *the homology groups have the structure of those of a finite complex*; (c) *at all points of the interval except at y itself every A can be shrunk away from its B.*

§2. Critical values and sets

6. **Definition.** The value y is said to be *critical* whenever it is impossible to chain-shrink A^y from B^y; it is *non-critical* otherwise.

THEOREM III. *A n.a.s.c. for y to be non-critical is that every c_p of A^y be chain-deformable onto U^y.*

The necessity of the condition follows from the definition. To prove its sufficiency we notice that as a consequence of it every c_p on A^y may be chain-deformed into a $c_p' \subset U^y$ on A^y. Since U^y may be chain-shrunk onto some $A^{y'}$, $y' < y$, on itself, c_p' is deformable onto $A^{y'}$ and hence the same holds for c_p. Therefore A^y may be shrunk away from B^y, and y is non-critical.

COROLLARY. *A n.a.s.c. for y to be critical is that A^y contain some chain which is not chain-deformable onto U^y on A^y.*

It follows immediately from Th. II (c) that every y is interior to an interval in which it is the only possible critical value, or

THEOREM IV. *The critical values are isolated.*

Now let y_1, y_2 be two consecutive critical values. There exist two values y', y'', where $y_1 < y' < y'' < y_2$, such that $A^{y'}$ may be chain-shrunk onto A^{y_1}, and U^{y_2} onto $A^{y''}$ away from B^{y_2}. Moreover, every point y^* of the closed interval (y', y'') is interior to an interval y_3y_4 such that A^{y_4} may be chain-shrunk onto A^{y_3}. By the Borel theorem, the closed interval $(y'y'')$ may be covered with a finite number of intervals such as y_3y_4. Hence by the product property of chain-deformation (DJ, p. 7) we have

THEOREM V. *The set U^{y_2} may be chain-shrunk onto A^{y_1}.*

By applying Theorem II (a) to the successive overlapping intervals considered, we have

THEOREM VI. *The homology groups of the sets A^y, U^y are fixed and the same throughout an open interval y_1, y_2 free from critical points, and they are those of A^{y_1} and also of U^{y_2}. In particular A^{y_1} and U^{y_2} have the same homology groups.*

7. **Characterization of the critical values by relative homology properties.** Let y be critical and c_p a chain of A^y not chain-deformable onto U^y. Setting $\gamma_{p-1} = F(c_p)$, we have two alternatives. The first is that γ_{p-1} is itself chain-deformable into a γ_{p-1}' of U^y on A^y. Then

$$A^y \supset d_p \to \gamma_{p-1}' - \gamma_{p-1}\,; \qquad c_p + d_p \to \gamma_{p-1}' \subset U^y.$$

In that case $\delta_p = c_p + d_p$ is a cycle of A^y mod U^y, and I say that $\delta_p \not\sim 0$ mod U^y on A^y. For in the contrary case

$$A^y \supset d_{p+1} \to c'_p - \delta_p = c'_p - c_p - d_p, \qquad c'_p \subset U^y.$$

This shows that c_p is chain-deformable into c'_p of U^y, with d_{p+1}, d_p as the deformation-chains. As this contradicts the hypothesis, our assertion follows. In the case considered, then, the p^{th} homology group of A^y mod U^y is not zero.

On the other hand, when γ_{p-1} is not chain-deformable onto U^y, it is $\not\sim 0$ mod U^y on A^y. For the contrary implies that

$$A^y \supset d_p \to \gamma'_{p-1} - \gamma_{p-1}, \qquad \gamma'_{p-1} \subset U^y,$$

and hence γ_{p-1} is chain-deformable onto U^y. Therefore here the $(p - 1)^{\text{st}}$ homology group of A^y mod U^y is $\neq 0$.

Thus if y is critical the homology groups of A^y mod U^y are not all zero. Conversely, when this holds, there exists a cycle c_p of A^y mod U^y, $\not\sim 0$, and hence not chain-deformable onto U^y. For in any case, since $\gamma_{p-1} = F(c_p) \subset U^y$ some $A^{y'} \supset \gamma_{p-1}$, $y' < y$. Then chain-deformability of the type indicated could be carried out so as to leave untouched γ_{p-1}. Therefore we should have a deformation with $\mathfrak{D}\gamma_{p-1} = 0$, and hence

$$A^y \supset \mathfrak{D}c_p \to c'_p - c_p, \qquad c'_p \subset U^y,$$

and consequently $c_p \sim 0$ mod U^y on A^y. We have therefore

THEOREM VII. *A n.a.s.c. for y to be critical is that some homology group of A^y mod U^y be $\neq 0$.*

Remark. It appears highly probable that even in the simple case of Ex. I of No. 2, the critical values depend upon the type and coefficient-group of the chains. Thus it is conceivable that with singular chains y may be critical, say mod m, but not in the system of integer coefficients. It would be interesting to prove this by examples.

8. **Critical sets.** When y is critical, A^y cannot be chain-shrunk onto U^y at every point of B^y, or it would be chain-shrinkable away from B^y (DJ, Th. III). A point x of B^y from which A^y cannot be chain-shrunk onto U^y is called *critical*, and the totality G^y of critical points on B^y is the *critical set* corresponding to y. The definition of critical point implies that if x is non-critical, there is a relatively open set $V \supset x$ on A^y whose chains are all deformable onto U^y on $U^y + V$. In the strict sense, $U^y + V$ is chain-shrinkable from B^y onto U^y. It follows from this that $V \cdot B^y$ consists of non-critical points. Therefore the non-critical points of B^y make up a relatively open set of B^y, and consequently the critical set G^y is *closed*. We shall now prove the important

THEOREM VIII. *The homology groups of A^y mod U^y (critical homology groups), of A^y mod $A^y - G^y$, and of $U^y + G^y$ mod U^y are the same.*

The proof will consist in showing that the first two are identical with the third. The equality of the first two means in substance that the critical groups are the so-called groups of the critical sets in A^y. These are the groups in which the chains c_p such that $F(c_p)$ does not meet G^y are the cycles, while ~ 0 means that $C_p \sim$ a chain not meeting G^y.

9. Let $V^1, V^2, \cdots$, be a decreasing sequence of open sets ($\overline{V^{i+1}} \subset V^i$) of A^y converging on G^y, and let

$$F^i = \overline{(V^{i-1} - V^i)} \cdot B^y, \qquad (V^0 = A^y). \tag{9.1}$$

Since A^y may be chain-shrunk away from every point of F^1 and F^1 is compact, it may be chain-shrunk away from F^1. Hence there is an open set $W^1 \supset F^1$ such that every chain of A^y may be chain-deformed to the exterior of W^1.

Now let c_p be a cycle of A^y mod U^y so that $\gamma_{p-1} = F(c_p) \subset U^y$ or else $\equiv 0$. In any case, since $F^1 \subset B^y$, we have $F^1 \cdot U^y = 0$ and hence $F^1 \cdot |\gamma_{p-1}| = 0$. Since $|\gamma_{p-1}|$ is closed, there exists an open set W' such that $F^1 \subset W' \subset W^1$, $W' \cdot |\gamma_{p-1}| = 0$. Since we may replace W^1 by W', we may assume W^1 so chosen that $W^1 \cdot |\gamma_{p-1}| = 0$. Now the chain-deformation of c_p may be so chosen that its part outside W^1 remains fixed. Therefore

$$\overline{W^1} \supset d^1_{p+1} \to c^1_p - c_p, \qquad c^1_p \subset V^1. \tag{9.2}$$

The same reasoning may be repeated with c^1_p, V^2, F^2, etc. We thus obtain for every $i > 1$,

$$\overline{W^i} \supset d^i_{p+1} \to c^i_p - c^{i-1}_p, \qquad c^i_p \subset V^i, \tag{9.3}$$

where W^i is an open set $\supset F^i$, but $\not\supset F(c^{i-1}_p) = \gamma_{p-1}$. Furthermore, since in the sequence $\{F^i\}$ only consecutive sets meet, we may choose the W's so that in the sequence $\{W^i\}$ likewise only consecutive sets meet. As a consequence in the sequence $\{d^i_{p+1}\}$ again only consecutive sets can meet.

10. Now let K be a fundamental complex for A^y and in K let C^i_{p+1}, C^i_{p+2} be the images of c^i_p, d^i_{p+1}, and let ϕ^h denote the sections of K (the skeleton of the coverings serving to construct K). Since only consecutive d's intersect, there exists an h_1 such that above ϕ^{h_1}, D^1_{p+2} is met by no chain D^i_{p+2}, $i > h_1$. We cut short all the D's by removing the part below ϕ^{h_1}. We proceed similarly for every D^i_{p+2} and the chains D^j_{p+2}, $j > i + 1$, so that the chains from D^i_{p+2} on are cut short at ϕ^{h_i}, where of course $h_i \geqq h_{i-1}$, and in any case we shall choose the inequality sign so that $h_i \to \infty$ with i.

Since only consecutive D's, as now modified, intersect,

$$D_{p+2} = D^1_{p+2} + D^2_{p+2} + \cdots, \tag{10.1}$$

$$D^{(h)}_{p+2} = D^{h+1}_{p+2} + \cdots \tag{10.2}$$

are both true chains of K and

$$D = D^1 + \cdots + D^h + D^{(h)}.$$

As we have frequently done on previous occasions, let us assume the whole configuration immersed in the Hilbert parallelotope $\mathfrak{H}$. Since D^i_{p+2} is above ϕ^{h_i} which $\to B^y$ with increasing i, and $|d^i_{p+1}| = \overline{|D^i_{p+2}|} \cdot B^y \to G^y$, necessarily

$|D^i_{p+2}| \to G^y$. Therefore the finite part of D^i_{p+2} tends also to G^y. A similar remark applies to $D^{(h)}_{p+2}$, since

$$(10.3) \qquad | D^{(h)}_{p+2} | \subset | D^{h+1}_{p+2} | + \cdots .$$

Now (9.2) and (9.3) imply

$$(10.4) \qquad D^i_{p+2} \to C^i_{p+1} - C^{i-1}_{p+1} + (\cdots), \qquad (C^0_{p+1} = C_{p+1}),$$

where $(\cdots)$ stands for some finite chain that need not be specified. Therefore,

$$(10.5) \qquad D_{p+2} \to C^h_{p+1} - C_{p+1} + F(D^{(h)}_{p+2}) + (\cdots).$$

From this follows immediately that if we choose a fixed k, we may take h so large that the infinite part of $F(D_{p+2})$ outside of V^k is the same as for $- C_{p+1}$. Therefore if we set

$$(10.6) \qquad D_{p+2} \to C'_{p+1} - C_{p+1},$$

we have $| C'_{p+1} | \cdot B^y \subset V^k$ for every k, and hence $\subset \Pi V^k = G^y$. Therefore C'_{p+1} represents a chain c'_p of A^y which meets B^y in points of G^y only.

Now the chain D_{p+2} represents a d_{p+1} of A^y such that, owing to (10.6),

$$d_{p+1} \to c'_p - c_p.$$

Hence $c'_p \sim c_p$ on A^y, and *a fortiori* mod U^y. That is to say, c_p, considered as a cycle of A^y mod U^y, is homologous to a cycle of $U^y + G^y$ mod U^y.

Suppose now that $c'_p \sim 0$ on A^y mod U^y, or

$$A^y \supset c_{p+1} \to c'_p + c^*_p, \qquad c^*_p \subset U^y.$$

Now all the chain-deformations considered in the previous discussion occurred in $A^y - G^y$. Since $| F(c_{p+1})| \cdot B^y \subset G^y$, we may apply all we have said to c_{p+1} and thus reduce it to a $c'_{p+1} \subset U^y + G^y$ without changing its boundary. It follows that if $c'_p \sim 0$ mod U^y on A^y, it is ~ 0 likewise on $U^y + G^y$. From this and linearity in the correspondence $c_p \leftrightarrow c'_p$ follows that the homology groups of A^y mod U^y and of $U^y + G^y$ mod U^y are the same.

Take now for c_p a cycle of A^y mod $A^y - G^y$, and set again $\gamma_{p-1} = F(c_p)$. We may choose γ so that the set $E = | \gamma_{p-1} | \cdot B^y$ is closed and contains no critical point, and hence γ_{p-1} may be deformed away from E onto U^y on some open set $U^y + V$ such that $V \supset E$, $\bar{V} \cdot G = 0$. Moreover, the chain-deformation may be confined to $\bar{V}$. Therefore there is a

$$d_p \to \gamma^*_{p-1} - \gamma_{p-1}; \qquad \gamma^*_{p-1} \subset U^y; \qquad | d_p | \cdot G^y = 0.$$

Consequently

$$c^*_p = c_p + d_p \sim c_p \text{ mod } A^y - G^y \text{ on } A^y, \qquad F(c^*_p) \subset U^y.$$

Now treating c^*_p as we did c_p we reduce it to a chain c'_p on $U^y + G^y$, $\sim c^*_p$ mod $(A^y - G^y)$. If $c'_p \sim 0$ mod $A^y - G^y$ on A^y we treat the chain which it bounds

as in the preceding case, and find once more that the homology groups are the same mod $A^y - G^y$ on A^y as mod U^y on $U^y + G^y$. This proves our theorem.

COROLLARY. *The critical homology groups for y are topological simultaneous invariants of G^y and of its neighborhood in A^y. Furthermore, if G^y consists of a finite number of isolated disconnected sets G^{iy}, this also holds as regards each connected component G^{iy}.*

The latter part follows from the fact that the groups for G^y are the direct sums of those for the sets G^{iy}.

11. **Isolated critical points.** If G_y^* is an isolated point-component of G^y, the corresponding homology groups are the *groups of the point* G^* in A^y. Suppose for example that A^y is locally polyhedral in G^*. We can construct for G^* in A^y a neighborhood which is the star $\mathrm{St}(G^*)$ of G^* in a simplicial complex K with the point as vertex. The sum of the faces of the cells of $\mathrm{St}(G^*)$ which do not have G^* as a vertex ("*complement*" or *linked-complex* of G^* in K) is a closed subcomplex L of K. The zero-cycles of G are $\not\sim 0$ when and only when G^* is an isolated point of A^y, and then the group of G^* for the dimension zero is the cyclic group generated by G^*. This corresponds to the case where a certain component of $\mathfrak{R}$ is "tangent" to y = constant in G^*, i.e., appears for the first time in B^y but not below.

When the preceding circumstance does not arise, the p^{th} homology group $(p > 1)$ of G^* is isomorphic to the $(p-1)^{\text{st}}$ of L, while for $p = 1$ it is isomorphic to the zeroth group of L mod one point of L. The proof is as follows. (a) By the fundamental deformation theorem, *Topology*, p. 86, the homology groups are shown to be the same as those corresponding to subchains of subdivisions of K and hence to $\mathrm{St}(G^*)$ in K; (b) the reduction to L follows then from the formula (9′) (loc. cit., p. 111) for joins.

Important remark. The reduction of the groups to those of L provides in the present case a systematic method for computing the groups and their homology characters (Betti numbers, type numbers, etc.).

Isolated critical sets. When the set G^y is HLC its Betti number R_0 is finite, and hence G^y consists of a finite number of "well-chained" compact sets G^{iy}. In fact this holds already when G^y is 0-HLC. When it is actually 0-LC (arcwise locally connected), every G^{iy} is connected, and G^y consists of a finite number of disjointed connected compact sets.

Important special case. The critical sets are all locally polyhedral. Then each consists of a finite number of disjointed connected compact sets.[6]

§3. Morse's type numbers (critical Betti numbers)

12. When the domain $\mathfrak{M}$ of the coefficients is a field, we may introduce the Betti numbers of the homology groups. In particular, Th. II b implies that the *Betti numbers* of A^y and U^y are all finite.

[6] See DJ, Th. XII, also Th. IV of S. Lefschetz, *On locally connected and related sets*, Annals of Mathematics, vol. 35 (1934), pp. 118–129.

The Betti numbers $R_p(A^y, U^y)$ are the type numbers of Morse for the value y.[7] They may also be called *critical Betti numbers*, for according to Th. III they are all necessarily zero when y is non-critical, so that *a sufficient condition for y to be critical* is that a corresponding critical Betti number be $\neq 0$. In the applications this is perhaps the chief property of these numbers. We have here considerable analogy with the Kronecker index of two cycles on a manifold, whose non-vanishing insures that they intersect. Similarly, also, when the number of signed coincidences or fixed points in transformations $\neq 0$, we know that coincidences or fixed points do exist.

Taking now a critical value y we may expect the presence of p-cycles of A^y mod U^y which are $\not\sim 0$ mod U^y. To have a base[8] for these cycles, I say that we merely need to take the following two sets:

(a) a maximal set γ_p^i $(i = 1, 2, \cdots, r_p)$ of absolute cycles of A^y independent mod U^y;

(b) a set of cycles $\delta_p^i \to \gamma_{p-1}^{*i}$, where γ_{p-1}^{*i} $(i = 1, 2, \cdots, s_{p-1})$ is a maximal independent set of cycles of U^y consisting of cycles ~ 0 on A^y.

r cannot exceed the maximum number of independent p-cycles of A^y, nor s the same number for the $(p-1)$-cycles of U^y:

$$r \leqq R_p(A^y), \; s \leqq R_{p-1}(U^y),$$

and hence r, s, like the Betti numbers themselves, are both *finite*.

To prove the base property we must show that:

I. Every cycle Γ_p of A^y mod U^y is a linear combination of γ's and δ's.

II. The γ's and δ's are independent.

If $\Gamma_{p-1} = F(\Gamma_p)$, we have by the maximal property of the cycles γ_{p-1}^*

$$U^y \supset c_p \to \Sigma\, y_i \gamma_{p-1}^{*i} - \Gamma_{p-1}, \tag{12.1}$$

$$\Gamma_p + c_p - \Sigma\, y_i \delta_p^i \to 0. \tag{12.2}$$

Since the left hand side is an absolute cycle of A^y, and $c_p \subset U^y$, we have by the maximal property of the γ_p's

$$\Gamma_p \sim \Sigma\, x_i \gamma_p^i + \Sigma\, y_i \delta_p^i \text{ mod } U^y \text{ on } A^y. \tag{12.3}$$

If the γ's and δ's were not independent, we could have in (12.3) $\Gamma_p = 0$, and yet not every x_i, y_i zero. Then

$$A^y \supset c_{p+1} \to \Sigma\, x_i \gamma_p^i + \Sigma\, y_i \delta_p^i - c_p, \qquad c_p \subset U^y. \tag{12.4}$$

[7] This definition has already been exploited by A. B. Brown (loc. cit.) on our suggestion. It is equivalent to that of Morse but topologically is simpler. Morse defines the type numbers as sums of two Betti numbers, neither of which is a topological character of the critical set, whereas under our definition the type numbers have an explicit topological character.

[8] Merely a maximal independent set, since the domain $\mathfrak{M}$ of the coefficients of the chains is a field.

The boundary of c_{p+1} being an absolute cycle, its own boundary $\equiv 0$. Hence

$$U^y \supset F(c_p) \equiv \Sigma\, y_i F(\delta_p^i) = \Sigma\, y_i \gamma_{p-1}^{*i} \sim 0 \text{ on } U^y. \tag{12.5}$$

Owing to the maximal property of the cycles δ_p, this implies that every $y_i = 0$. But then (12.4) would imply that the γ's are dependent mod U^y, which is contrary to assumption. Therefore every $x_i = 0$, which proves II, and the γ's and δ's together do form a base.

Since the Betti number is equal to the number of elements in the base, we have for the p^{th} type number corresponding to y

$$R_p(A^y, U^y) = r_p + s_{p-1} \leqq R_p(A^y) + R_{p-1}(U^y). \tag{12.6}$$

Since the right hand side is clearly finite, we have

THEOREM IX. *The type numbers for any critical value are all finite.*

Furthermore from No. 11 Corollary follows

THEOREM X. *The type numbers are topological simultaneous invariants of the critical set G^y and of its neighborhood in A^y.*

13. Modifying our notation somewhat, let us write

$$\begin{aligned} r_p &= \Delta' R_p(U^y), \qquad s_p = \Delta'' R_p(U^y), \\ \Delta R_p(U^y) &= R_p(A^y) - R_p(U^y). \end{aligned} \tag{13.1}$$

The notation is justified on the following ground. In the first place, if ϵ is less than the distance from y to the next higher critical value, then $R_p(U^{y+\epsilon}) = R_p(A^y)$, and hence

$$\Delta R_p(U^y) = R_p(U^{y+\epsilon}) - R_p(U^y). \tag{13.2}$$

In the second place, I say that

$$\Delta R_p(U^y) = (\Delta' - \Delta'')\, R_p(U^y) = r_p - s_p. \tag{13.3}$$

Consider in fact the following cycles: the maximal sets γ_p^i, γ_p^{*i} of No. 12, and a new maximal set γ_p^{**i} $(i = 1, 2, \cdots, t_p)$ of p-cycles of U^y independent on A^y.

Here again $t_p \leqq R_p(U^y)$; hence t_p is finite.

I now say that

(a) the sets γ_p, γ_p^{**} together form a base for the absolute p-cycles of A^y;

(b) the sets γ_p^*, γ_p^{**} together form a base for those of U^y.

In each case the definitions of the sets imply that the expected bases consist of elements independent in the appropriate sense.

Let Γ_p be an absolute cycle of A^y. From the definition of the γ's and γ^{**}'s we have in succession on A^y

$$\begin{cases} \Gamma_p \sim \Sigma\, x_i \gamma_p^i + \Gamma_p', & \Gamma_p' \subset U^y; \\ \Gamma_p' \sim \Sigma\, y_i \gamma_p^{**i}; & \\ \Gamma_p \sim \Sigma\, x_i \gamma_p^i + \Sigma\, y_i \gamma_p^{**i}. & \end{cases} \tag{13.4}$$

This proves (a). Similarly if Γ_p is a cycle of U^y, we have on U^y

$$(13.5)\qquad \begin{cases} \Gamma_p \sim \Sigma x_i \gamma_p^{**i} + \Gamma'_p, & \Gamma'_p \sim 0 \text{ on } A^y; \\ \Gamma'_p \sim \Sigma y_i \gamma_p^{*i}; \\ \Gamma_p \sim \Sigma x_i \gamma_p^{**i} + \Sigma y_i \gamma_p^{*i}. \end{cases}$$

This proves (b). From (a) and (b) follow

$$(13.6)\qquad R_p(A^y) = t_p + \Delta' R_p(U^y),$$

$$(13.7)\qquad R_p(U^y) = t_p + \Delta'' R_p(U^y),$$

from which (13.3) follows.

14. Let us designate by ΔM_p the p^{th} type number for y, so that by (12.6)

$$(14.1)\qquad \Delta M_p = R_p(A^y, U^y) = \Delta' R_p(U^y) + \Delta'' R_{p-1}(U^y).$$

Combining (14.1) and (13.3) we find

$$(14.2)\qquad \Delta(M_p - R_p(U^y)) = \Delta''(R_{p-1}(U^y) + R_p(U^y)).$$

This holds also for $p = 0$, provided that we write $R_{-1} = \Delta'' R_{-1} = 0$. From (14.2) we conclude immediately by summation

$$(14.3)\qquad (-1)^k \Delta \sum_{p=0}^{k} (-1)^k (M_p - R_p(U^y)) = \Delta'' R_k(U^y) = s_k \geqq 0,$$

and directly

$$(14.4)\qquad \Delta(M_p - R_p(U^y)) \geqq 0.$$

These are the basic "local" inequalities for the type numbers due to Morse.[9]

Let us write

$$(14.5)\qquad M_p^{a,b} = \sum_{a \leqq y \leqq b} \Delta M_p,$$

$$(14.6)\qquad R_p^{a,b} = \sum_{a \leqq y \leqq b} \Delta R_p(U^y) = R_p(A^b) - R_p(U^a).$$

Then we have by summation the actual inequalities of Morse:

$$(14.7)\qquad (-1)^k \sum_{p=0}^{k} (-1)^p (M_p^{a,b} - R_p^{a,b}) \geqq 0,$$

$$(14.8)\qquad M_p^{a,b} \geqq R_p^{a,b}.$$

We must now distinguish two cases according as the range of $f(x)$ is closed or open upwards, i.e., according as $\mathfrak{R}$ is compact or not.

[9] For all Morse's relations see MC, Chapter VI.

15. **Case I. The space $\mathfrak{R}$ is compact.** Then the range of y is finite and closed, say the closed interval a, b. We have then

$$R_p^{a,b} = R_p(A^b) = R_p,$$

the Betti number of the space itself. Designating by M_p the p^{th} *type number* of $\mathfrak{R}$ itself, or sum of the numbers ΔM_p for all the critical values, we shall have $M_p^{a,b} = M_p$. Finally, since $\mathfrak{R}$ is HLC, the Betti numbers, and the numbers $\Delta'' R_k(U^y)$ for k above a certain fixed integer n are all zero. In particular, if dim $\mathfrak{R}$ is finite, we may take $n = \dim \mathfrak{R}$. Under these circumstances (14.7) and (14.8) yield

$$(15.1) \qquad (-1)^k \sum_{p=0}^{k} (-1)^p (M_p - R_p) \geqq 0, \qquad k < n,$$

$$(15.2) \qquad \sum_{p=0}^{n} (-1)^p M_p = \sum_{p=0}^{n} (-1)^p R_p,$$

$$(15.3) \qquad M_p \geqq R_p, \qquad 0 \leqq p \leqq n.$$

We obtain (15.2) from (14.3) by taking $k = n + 1$ and observing that here $M_{n+1} = R_{n+1} = \Delta'' R_{n+1}(U^y) = 0$. The last three relations are the basic equality and inequalities of Morse for the whole of the space $\mathfrak{R}$.

Remarks. I. We have considered the numbers ΔM_p as the type numbers corresponding to y. We have, however, by Th. VIII, if G^y is the critical set for y,

$$\Delta M_p = R_p(A^y, U^y) = R_p(A^y, A^y - G^y).$$

Hence ΔM_p might equally well be called the p^{th} type number of the critical set. Furthermore, suppose that G^y consists of a finite number of disjointed connected sets, G^{iy}. Since they are all closed, we have at once

$$R_p(A^y, A^y - G^y) = \Sigma R_p(A^y, A^y - G^{iy}).$$

The numbers in the sum are the p^{th} type numbers of the individual critical sets. We might denote them by $\Delta^i M_p$ and we should have again

$$M_p^{ab} = \sum_{a \leqq y \leqq b} \Delta^i M_p, \qquad M_p = \sum \Delta^i M_p.$$

Therefore, in (14.7), $\cdots$, (15.3) the M's may also be considered as the sum of the type numbers for the isolated critical sets.

II. Theorem X holds for the isolated critical sets G^{iy}.

III. While the local type numbers depend solely upon a neighborhood of B^y in A^y, this is not the case as regards the local numbers $\Delta' R$, $\Delta'' R$. Consider one of the absolute cycles γ_p^i, of which there are $\Delta' R_p$. It may very well happen that, say, γ_p^1 is homologous on $\mathfrak{R}$ to a cycle in the region $A^y - U^a$, but not in any similar region corresponding to any value higher than a. If we assume then that there is removed from $\mathfrak{R}$ a set $U^{a'}$, $a < a' < y$, there will cease to be

a cycle $\sim \gamma_p^1$ on the *new* A^y. The result is readily seen, in fact, to be a decrease in the number of cycles of the base $\{\gamma_p^i\}$ with an equal increase in those of the base $\{\gamma_p^{*i}\}$, thus leaving fixed ΔM_p, which is strictly local in its nature.

16. **Case II.** $\mathfrak{R}$ **is merely locally compact.** That is to say, the range of y is open above, $a \leqq y < b$. Our treatment is applicable throughout any closed subinterval of the range, and even to the whole range if the number of critical values is finite.

Suppose, however, that the critical values form an infinite sequence $y_1, y_2, \cdots$, necessarily $\rightarrow b$. Let us write

$$M_p^h = M_p^{a, y_h}, \qquad R_p^h = R_p^{a, y_h},$$

and let M_p, R_p be as before. In any case

$$M_p^h \geqq R_p^h, \tag{16.1}$$

$$M_p = \lim_{h \to \infty} M_p^h. \tag{16.2}$$

Now $R_p = \infty$ means the maximum number of independent cycles on $A^{y_h} \rightarrow \infty$ with h, and hence $R_p^h \rightarrow \infty$, $M_p^h \rightarrow \infty$, $M_p = \infty$. It follows that if M_p is finite the same holds for R_p.

On the other hand, let R_p be finite and let γ_p^i $(i = 1, 2, \cdots, R_p)$ be a maximal independent set for the p-cycles of $\mathfrak{R}$. Then for h above a certain value, the γ's will all be on A^h and will be necessarily independent. Therefore $R_p^h \geqq R_p$ and (15.3) follows.

Suppose now that *every* M_p, $p \leqq q + 1$, *is finite*. Then I say that (15.1) holds for every $k \leqq q$. We observe first that R_p is finite for every $p \leqq q + 1$. Furthermore, from (16.2) we learn that when h exceeds a certain limit r, $M_p^h = M_p$ for every $p \leqq q + 1$. Therefore for sufficiently high h the p^{th} type numbers, $p \leqq q + 1$, for y_h will all be zero.

Let us then choose h so high that the preceding conditions already hold for h, and that for every $p \leqq q$, A^{y_h} contains a maximal independent set of p-cycles γ_p^i $(i = 1, 2, \cdots, R_p)$ of the space $\mathfrak{R}$. I say that $R_p^h = R_p$. For let Γ_p be any p-cycle of A^{y_h}. We have on $\mathfrak{R}$, and hence on some A^{y_j}, $j \geqq h$,

$$c_{p+1} \rightarrow \Gamma_p - \Sigma x_i \gamma_p^i. \tag{16.3}$$

Let us suppose $j > h$. Since $F(c_{p+1}) \subset A^{y_h} \subset U^{y_j}$, c_{p+1} is a cycle of A^{y_j} mod U^{y_j}, and since the type number of index $p + 1 \leqq q + 1$, for the critical value $y_j > y_h$, is zero,

$$c_{p+1} \sim 0 \text{ mod } U^{y_j} \text{ on } A^{y_j}.$$

Therefore c_{p+1} is deformable onto U^{y_j} without changing its boundary, hence also onto $A^{y_{j-1}}$, and consequently step by step onto A^{y_h}. Therefore (16.3) holds on the latter and the p-cycles of A^{y_h} depend on the cycles γ_p^i. It follows that $R_p^h \leqq R_p$, and hence both are equal.

Now (14.7) holds with M_p^{ab}, R_p^{ab} replaced by $M_p^h = M_p$, $R_p^h = R_p$, for every $p \leqq q$, so that (15.1) holds as asserted. That is to say, *if the type numbers M_p, $p \leqq q + 1$, are all finite, this holds also for the numbers R_p of same index and* (15.1) *holds for every $k \leqq q$.*

§4. Various extensions

17. We shall consider two types of extensions. The first will refer to the structure of the sets A^y, the second to the homology groups themselves. *All the results obtained so far hold in both cases,* and this is true notably regarding the type number formulas.

Extension to a wider class of sets A^y. While there is considerable scope in this direction, it will be sufficient for the applications to replace Axiom I by the weaker

Axiom I′. *Every A^y is chain-shrinkable onto a compact subset A^{*y} obeying Axiom I and such that $A^{*y} \supset A^{*y'}$ for $y > y'$.*

In other words, A^{*y} is compact HLC and all the sets $A^{y'} \cdot A^{*y}$ are also HLC. We still preserve Axiom II.

We now examine the extension of our theorems. By virtue of the Lemma of No. 3, Theorem I follows.

Given any $\epsilon > 0$ consider $A^{*y+\epsilon}$. Since Axiom I holds for $A^{*y+\epsilon}$, there exists an η such that every c_p of $A^{y+\eta} \cdot A^{*y+\epsilon}$ is ϵ chain-deformable onto $A^y \cdot A^{*y+\epsilon}$, and hence by Axiom I′, chain-deformable onto A^{*y}. This is all that we need for the reasoning of No. 4. Now if $c_p \subset U^y$, we chain-deform it first into a chain c_p' of A^{*y}. Since $B^y \cdot A^{*y}$ is compact, we may (Axiom II and DJ, Theorem III) chain-deform c_p' over A^y onto a set $A^{y'}$, $y' < y$, and then over $A^{y'}$ onto $A^{*y'}$. Therefore c_p may be chain-shrunk away from B^y onto a fixed $A^{*y'}$, $y' < y$ over A^y. Together with the preceding result this suffices to prove Theorem II. The rest of the treatment is then exactly as before. The only additional remark to be made is that since A^y may be chain-shrunk onto A^{*y}, the critical set G^y is on $A^{*y} \cdot B^y$, as follows at once from its definition (No. 8). In particular, G^y is still compact here.

18. **Extension to different types of homology groups.** If $\mathfrak{S}$ is a subset of $\mathfrak{R}$ we may consider throughout in place of homology groups absolute, mod U^y, etc., the following two types:

(α) the relative homology groups of the similar types mod $\mathfrak{S}$;

(β) the absolute homology groups taken as if $\mathfrak{R} - \mathfrak{S}$ were the basic space. That is to say, A^y, U^y, etc., are to be replaced by $A^y - \mathfrak{S}$, $U^y - \mathfrak{S}$ throughout.

We shall refer to these two cases respectively as type, critical sets, etc., of $\mathfrak{R}$ mod $\mathfrak{S}$, and of $\mathfrak{R} - \mathfrak{S}$. Each requires a suitable modification of the axioms.

Type $\mathfrak{R} - \mathfrak{S}$. We impose in addition to Axioms I, II the new

Axiom III. *The sets $A^y \cdot \mathfrak{S}$ are compact.*

Axiom IV. *Every $A^y - \mathfrak{S}$ may be chain-shrunk away from $\mathfrak{S}$ over itself.*

Our whole treatment applies here with scarcely a modification. In particular,

every chain of $A^y - \mathfrak{S}$ is chain-deformable over $A^y - \mathfrak{S}$ into a subchain of a finite chain-complex. Hence, again its homology structure is that of a finite complex, the Betti numbers are all finite and so are the type numbers. They all refer of course to $\mathfrak{R} - \mathfrak{S}$. The critical sets are still compact, and $\subset \mathfrak{R} - \mathfrak{S}$.

Type $\mathfrak{R}$ mod $\mathfrak{S}$. We now impose Axioms I, II, III, but in place of IV we choose

AXIOM IV′. *The sets* $A^y \cdot \mathfrak{S}$ *are* HLC.

The chain-deformations must be modified throughout in that they are to be taken mod $\mathfrak{S}$. With this understanding, and referring to the remarks of DJ No. 17, we find that our whole theory is directly applicable here also.

Examples of type $\mathfrak{R} - \mathfrak{S}$. I. The set $\mathfrak{R}$ is as in Example I, No. 2, and $\mathfrak{S}$ is a similar subset of $\mathfrak{R}$. Axiom III clearly holds. Taking now as the L of No. 2 the set $A^y \cdot \mathfrak{S}$ and operating as in the place cited, we find that Axiom IV holds also.

II. $\mathfrak{R}$ is as in Ex. IV of No. 2, and $\mathfrak{S}$ is a similar subset of $\mathfrak{R}$ such that the sets $A^y \cdot \mathfrak{S}$ are compact. Since $A^y \cdot \overline{(A^y - \mathfrak{S})}$ is locally polyhedral, the preceding reasoning is applicable at all points of $A^y \cdot \mathfrak{S}$, from which Axiom IV follows since $A^y \cdot \mathfrak{S}$ is compact. This case includes obvious analogous extensions of Examples II, III of No. 2.

Examples of type $\mathfrak{R}$ mod $\mathfrak{S}$. Axiom IV′ holds for the two examples just considered, and hence they constitute examples for this type also.

§5. Application to the critical points of functions

19. We have already formulated the problem (No. 2, Ex. IV, No. 18, Ex. II), and it falls under our general theory. However, when the function and the region are suitably restricted, the treatment may be carried somewhat further. In particular, an analytical characterization of the critical points becomes possible and, in the simpler cases, the type numbers are readily determined. The following conditions, while keeping down extraneous analytical difficulties, preserve sufficient generality and interest for most applications.

(a) $\mathfrak{R} = \Omega + \Phi$, where Ω is a connected region of a topological analytical $\mathcal{M}_n$, its boundary Φ being an analytical $(n - 1)$-variety on $\mathcal{M}_n$;

(b) the function f is analytical over $\mathfrak{R}$;

(c) the sets A^y: $f \leqq y$ are all compact.

Regarding any point P of Φ we have three possibilities. (a) P is singular, i.e., it has no analytical $(n - 1)$-cell E_{n-1} for neighborhood relative to Φ; (b) P has an E_{n-1} for neighborhood but the cell abuts on two disjointed n-cells both in Ω; (c) the same as (b), except that only one of the two n-cells is in Ω, the other being in $\mathcal{M}_n - \Omega$. In the first two cases we shall say that P is a *singular*, in the third that it is an *ordinary* point of the boundary.

Except for the generality of the boundary and analyticity in place of C^h, the situation is substantially the same as that considered by Morse. The treatment of the analytical case is readily adapted to C^r, $r \geqq 2$ or 3. The unrestricted boundary has been made possible in an effective way, because we utilize the

invariants of $\mathfrak{R} - \Phi$ or of $\mathfrak{R}$ mod Φ. We shall first investigate the local behavior of the type numbers, even determine them in a few simple cases, and conclude with the application of the general type number formulas.

20. **Critical points on the region** Ω. It will be more convenient to change our notation somewhat, designating a generic point of $\mathcal{M}_n$ by P, and an admissible coördinate system about P by $x_1, \cdots, x_n$. Occasionally we shall use a local metric over the range of validity of the x's, and it will then always be understood to be the euclidean metric for these coördinates.

If we have at P

$$f_{x_i} = 0 \qquad (i = 1, 2, \cdots, n), \tag{20.1}$$

this condition remains fulfilled in any other admissible coördinate system about P. When it holds, we shall call P an *analytical* critical point (= a.c.p.) by contrast with the type previously considered, or *topological* (= t.c.p.).

When P is an a.c.p., the rank r of

$$\| f_{x_i x_j} \|_P \tag{20.2}$$

and the number k of negative terms in the reduced form of the quadratic form

$$f_{x_i x_j} u_i u_j, \qquad x = P, \tag{20.3}$$

are also both independent of the coördinates chosen. If $r = n$, we call P a *non-degenerate* a.c.p., and k is its *index*. Such a point is necessarily isolated, for in that case the jacobian of (20.1), which is the determinant of (20.2), is $\neq 0$, and hence P is an isolated solution of (20.1).

21. We shall now prove simultaneously the following important propositions.

THEOREM XI. *The analytical and topological critical points on* Ω *coincide.*

THEOREM XII. *A non-degenerate analytical critical point of index* k *is a topological critical point whose type numbers are* $m_p = \delta_{pk}$, *the Kronecker delta, (Morse).*

It is important to bear in mind that both theorems refer to the region Ω but not to its boundary Φ.

First part of the proof of Theorem XI. We first show that if P is not an a.c.p., it is not a t.c.p. For under the assumption we may choose the coördinates with P as the origin and $f(x) = y = c + x_n$ so that about P the sets A^y and B^y are respectively represented by $x_n \leqq 0$, $x_n = 0$. A displacement downwards parallel to the x_n axis (assumed vertical) will then free a certain neighborhood of P in A^y from all its chains. Since the displacement about P is over A^y, it follows from DJ, Theorem I and the definition of the topological critical sets (No. 7) that P is not a point of such a set.

To prove Theorem XI, we must show conversely that a t.c.p. is also an a.c.p., but for this purpose we need Theorem XII which we shall therefore take up next.

22. Proof of Theorem XII.[10] It is sufficient to establish the type number property, for it implies that one of these numbers is $\neq 0$ and hence that P is a t.c.p. When P is a non-degenerate a.c.p., we may choose coördinates with P as origin and such that about P the expansion of f is of the form

$$(22.1)\qquad f(x) = c + \phi_2(x) - \phi_2'(x') + \cdots,$$

$$(22.2)\qquad \phi_2 = x_1^2 + \cdots + x_h^2, \qquad \phi_2' = x_1'^2 + \cdots + x_k'^2, \qquad h + k = n,$$

where the terms of order > 2 are omitted.

We propose to show that a suitable neighborhood of P on A^y may be homotopically deformed and hence chain-deformed over A^y onto its subset $\phi_2 = 0$, the homotopy leaving P invariant. It will follow that the critical homology groups are the same as those of P relative to the points with $x_i = 0, x_j'$ arbitrary, i.e., those of a point in an S_k. For such a point the (relative) Betti numbers are precisely δ_{pk}, so that the theorem will have been proved.

When $k = 0$, P is an isolated point of A^y and the required property is trivial, and when $k = n$ it is already fulfilled. Hence we may assume both $h, k \neq 0$.

We shall designate by S_h, S_k the two spaces $x_i' = 0$, $x_i = 0$. Let $Q(x, x')$ be any point near P and let q, q' be the corresponding points $(x, 0)$, $(0, x')$ of S_h and S_k, which are uniquely determined by Q. Without changing the form of (22.1), (22.2) or the metric about P, we may apply a linear change of coördinates such that the coördinates of Q, q, p' become

$$Q\colon x_1 = a, x_1' = a'; \qquad q\colon x_1 = a, x_1' = 0; \qquad q'\colon x_1 = 0, x_1' = a',$$

and all other coördinates zero, and if initially $d(P, Q) < \alpha$, then $|a|, |a'| < \alpha$.

Let π be the plane of the parallelogram $PqQq'$, and C its intersection with the locus $f = c$. The equations of the tangents to C at P referred to the axes in π are $x_1^2 - x_1'^2 = 0$. Since they are distinct and distinct from Pq, the curve has two branches through P not tangent to one another nor to Pq. Hence, for α small, $Qq'\colon x_1' = a'$ intersects C in two points, R, S, one on each branch, separated by q'. Now the sign of $f - c$ along Qq' changes only as R or S is crossed, and since it is minus in q', RS closed is the intersection of π with $A^y\colon f \leqq c$. Hence $Q \subset A^y$ implies: $Q \subset RS$ closed, segment $Qq' \subset A^y$. In other words, when $d(P, Q) < \alpha$, $Q \subset A^y$, Q may be joined in a unique way by a segment Qq' on A^y to a point q' on S^k not farther than α from P. The segments $\overrightarrow{Qq'}$ determine a homotopy, and hence a chain-deformation, of the α-neighborhood of P on A^y over itself, onto the same for P on S_k. As it has already been stated, this proves Theorem XII.

23. Completion of the proof of Theorem XI. The coördinates x being again chosen with P as origin, let us suppose that P is an a.c.p. but not a t.c.p.

[10] For Morse's proof see MC, p. 172. The one here given has been adapted to our definition of type numbers.

There is a spherical region ω of center P, $\subset \Omega$, and of radius ϵ so small that every $c_p \subset \omega$ may be chain-deformed to the exterior of ω onto U^y over A^y. For about P it may be deformed below a certain $y' < y$, and hence outside of a suitable ω which we take above y'.

Consider now the new function

$$F = f + u_i x_i + \frac{\lambda}{2} x_i x_i,$$

related to ω as f itself is to Ω. The hessian of F is

$$H = | F_{x_i x_j} | = \lambda^n + \cdots \not\equiv 0$$

in the variables λ, x. Therefore we may choose them such that it is $\neq 0$, and in particular choose a fixed λ, as we now assume done, such that $H \not\equiv 0$ in the variables x alone. Therefore the locus $H = 0$ will represent an analytical $(n - 1)$-variety about P. We may therefore choose a point Q on ${}_\omega \cdot U^y$ not situated on $H = 0$. If we now choose

$$u_i = - (f_{x_i} + \lambda x_i)_Q,$$

we find that Q satisfies the system

$$F_{x_i} = 0 \qquad (i = 1, 2, \cdots, n),$$

and hence Q is an a.c.p. for F, but since $H(Q) \neq 0$, it is a non-degenerate a.c.p. By Theorem XII Q is therefore a t.c.p. of F. Let A', U' be the sets $F \leqq F(Q)$, $F < F(Q)$ (the analogues of A^y, U^y through Q). We have shown that the chains of A' may be chain-deformed onto a subset which, except for Q, is in U' (the space S_k of No. 21). This holds in particular for a certain γ_p of A' mod U', which is $\not\sim 0$ mod U' on A'.

Since $Q \subset \omega \cdot U^y$ there is a chain-deformation ϑ over A^y reducing γ_p to a chain $\gamma_p' \subset A^{y'}$, and hence by DJ Theorem II, to a chain merely $\subset U^y \cdot U'_{-\omega} \subset U'$. Similarly for the two associated deformation-chains $\mathfrak{D}\gamma_p$, $\mathfrak{D}F(\gamma_p)$, which we may merely deform onto A' without moving their fixed part γ_p. As a consequence we may assume the $\mathfrak{D}$ chains on A', which means that we may assume that ϑ is over A', and hence $\gamma_p \sim 0$ mod U' on A'. This contradiction completes the proof of our theorem.

24. The device just used, of replacing f by a more convenient approximating function, occurs repeatedly in the work of Morse. It resembles the familiar device of algebraic geometry for replacing a complicated singularity, say of a curve, by an approximating simpler singularity. Geometrically, it is also the analogue of the method whereby we have defined intersections of cycles on a manifold by those of suitable approximating polyhedral cycles. Morse has made use of this device (MC, p. 175) in his computation of the number of non-degenerate critical points equivalent to a given critical set G^y. The scheme may be described thus. There is constructed an analytical function $F(x, \mu)$ such that

$F(x, 0) \equiv f$, and that within some interval $y - \epsilon, y + \epsilon$ it possesses only isolated non-degenerate critical points. The number m_p of those of type number p is equal to the p^{th} type number of G^y. The proof is immediate. For a suitable ϵ and small μ the sets $F \leqq y \pm \epsilon$ may be chain-deformed into the sets $A^{y \pm \epsilon}$ (for "between" the corresponding sets there will be found no critical points of F or of f), and similarly with f and F interchanged. It follows that the numbers ΔM_p corresponding to the interval $y - \epsilon, y + \epsilon$ are the same for both.

25. **Critical points on the boundary** Φ. The values of f on Φ are those of an analytical function g attached to Φ. We now have the basic

THEOREM XIII. *Let P be an ordinary point of Φ. If* (a) *P is not analytically critical for g or* (b) *P is a relative maximum for $f - g$ on $\mathfrak{R}$, then P is topologically non-critical for f on $\mathfrak{R}$.* (c) *If P is both analytically critical for g on Φ and a relative minimum for $f - g$ on $\mathfrak{R}$, it is topologically critical for f on $\mathfrak{R}$, and if P is isolated, its homology groups and type numbers as to $\mathfrak{R}$ are the same as those of P as to g on Φ.*

If P is not an a.c.p. of g on Φ, it is non-critical for f on $\mathfrak{R}$. For in that case we may choose a coördinate system about P such that

$$(25.1) \qquad \Phi: x_n = 0, \qquad \mathfrak{R}: x_n \geqq 0, \qquad f = c + x_{n-1},$$

so that A^y is the set $x_{n-1} \leqq 0$ about P. It follows that by a translation parallel to the x_{n-1} axis downwards, the chain c_p of A^y may be freed from a certain neighborhood of P and hence P is non-critical.

In case (b) we may choose the same representation for Φ and $\mathfrak{R}$, and now a translation parallel to the x_n axis upwards is a homotopy over A^y about P, shrinking its chains onto U^y away from P; hence P is not a t.c.p.

Consider finally case (c). At all events, by virtue of Theorem XI with Φ and g instead of Ω, f, we find that P is a t.c.p. for g on Φ. If A'^y is the analogue of A^y for $c + g$ on Φ, the projection of A^c parallel to the x_n axis onto Φ is an A'^c about P. Suppose now that P is not a t.c.p. for f on $\mathfrak{R}$. As a consequence, every $c_p \subset A'^c$ in a certain neighborhood V^p of P as to Φ may be chain-deformed over A^c onto a certain $A^{c-\epsilon}$. Applying the projection just mentioned to the associated deformation-chains we reduce them to Φ, and hence have in Φ a chain-deformation onto $A'^{c-\epsilon}$, which contradicts the fact that P is a t.c.p. for g on Φ. Therefore P is likewise a t.c.p. for f on $\mathfrak{R}$. Furthermore, if P is isolated, and since the projection of A^c onto A'^c is a chain deformation of A^c onto A'^c leaving P untouched, the statement as to the homology groups and type numbers follows. The proof of the theorem is thus completed.

COROLLARY I. *The topological critical points of f on $\mathfrak{R}$ are found among the singular points of Φ, and among the critical points of g where $f - g$ does not have a relative maximum on $\mathfrak{R}$.*

COROLLARY II. *If P is an ordinary point of Φ, which is non-degenerate critical of index k for g, and a relative minimum for $f - g$ on $\mathfrak{R}$, it is an isolated topological critical point for f on $\mathfrak{R}$ and its type numbers are $m_p = \delta_{pk}$.*

The last part of Corollary II follows from Theorems XII and XIII. This

corollary and Theorem XII embody the only simple cases where the computation of the type numbers may actually be carried through analytically.

Special case of Corollary II. The point P is not an a.c.p. for f so that in a suitable coördinate system $f = c + g + x_n$. In an arbitrary admissible coördinate system f has a positive normal derivative towards Ω. This case has been treated by Morse-Van Schaack.[11]

26. **Type number formulas.** Once we have obtained the type numbers, the general formulas (§3) of Morse become applicable. Three choices are open to us.

(a) The homologies and their invariants are those of $\mathfrak{R}$ itself. In that case we must include the type numbers of the critical points on the boundary also.

(b) The homologies are those of $\mathfrak{R} - \Phi$ or those of $\mathfrak{R}$ mod Φ. The boundary critical points may then be neglected. However, owing to the duality theorems for a relative $\mathcal{M}_n$ (*Topology*, Chapter 7), the two situations give rise to the same Betti numbers.

Applications. I. All the a.c.p. of f on Ω are non-degenerate. Then the basic formulas hold with M_p and the numbers R_p referring to Ω alone, but not to $\mathfrak{R}$.

II. The same as the preceding, but in addition Φ is a manifold $\mathcal{M}_{n-1}$, f has no a.c.p. on Φ and its normal derivative is increasing towards Ω at all points of Φ. Then the basic formulas hold as they stand. This is the actual situation considered by Morse in his earlier papers.

III. The same as II, except that g has non-degenerate critical points at which f is increasing towards Ω. Then the formulas hold as they stand, with M_p the number of critical points on Ω plus those of g on Φ where f is increasing towards Ω (Morse-Van Schaack, loc. cit.).

It is to be observed that even when one of the type number differences, of No. 14, say $M_p^{ab} \neq 0$ (a, b, non-critical), it does not follow that there are at least M_p^{ab} distinct critical values on ab, but merely that the sum of each with a suitable coefficient is M_p^{ab}. There may very well be just a single critical value taken M_p^{ab} times. There is here a clear analogy with the Kronecker index. Just as for the index, the function f may be replaced by a suitable approximating function F (No. 24) which will possess the requisite number M_p of isolated non-degenerate critical points of index p each counted once, whose y's will be on the interval ab. This may likewise be done over the boundary Φ as regards the critical points mentioned in Theorem XIII.

PRINCETON UNIVERSITY.

[11] See the last theorem of their paper in the Annals of Mathematics, vol. 35 (1934), pp. 545–571.

GENERALIZED MINIMAX PRINCIPLE IN THE CALCULUS OF VARIATIONS

By G. D. Birkhoff and M. R. Hestenes

Introduction

In the study of the critical points of a function $f(x_1, \cdots, x_n)$ one naturally begins with the maximum and minimum points. Similarly, the study of the critical extremals of an integral

$$J = \int_{t_1}^{t_2} f(x, \dot{x})\, dt$$

joining two fixed points begins with the properties of minimizing or maximizing extremals and the question of their existence. It was not until recently that a systematic study was made of critical points of functions and critical extremals of integrals which are not necessarily of the minimizing or maximizing type. This study seems to have had its beginning in 1917 in a paper by Birkhoff[1] in which he enunciated his minimax principle. Birkhoff treats only the critical points and critical extremals of the so-called type one. Beginning with a paper in 1925 Morse[2] has developed systematically by the use of Analysis Situs the existence and the relation between critical points and critical extremals of all types. A. B. Brown and a number of others have also written on this subject.[2]

The principal method used heretofore in obtaining the critical point relations is the following. A value b will be called a critical value of our functional $f(P)$ if there is a critical point Q of $f(P)$ such that $f(Q) = b$. We now consider the connectivities of the domains $f(P) \leqq b$, as the constant b varies from the absolute minimum b_0 of $f(P)$ on the domain under consideration. It is found that the connectivities of $f \leqq b$ change only when the variable b passes through a critical value of $f(P)$. The change in connectivity depends upon the type of the critical point P having this critical value. By studying these changes of connectivity one is able to classify the critical points of $f(P)$ and to obtain the critical point relations. This method was used by Birkhoff in order to obtain his minimax principle and by Morse to obtain a complete set of critical point

Received June 12, 1935.

[1] *Dynamical systems with two degrees of freedom*, Transactions of the American Mathematical Society, vol. 18 (1917), pp. 199–300.

[2] For references to literature on this subject see Morse, *Calculus of Variations in the Large*, Colloquium Lectures, American Mathematical Society, vol. 18 (1934). Unless otherwise expressly stated, all references to Morse are to his book. See also Morse and Van Schaack, *Abstract critical sets*, Proceedings of the National Academy of Sciences, vol. 21 (1935), pp. 258–62.

relations. Lately Lefschetz has announced some further results, apparently using this method.[3]

In the above method, the notion of minimum and maximum are seen to be special instances of a more general notion of the type of a critical point. The question naturally arises: cannot this process be reversed? Cannot all these critical relations be obtained by a minimizing principle? We propose to show in this paper that this can be done.[4] Our primary notion is essentially that of finding the minimum of the maximum $F(C_k)$ of $f(P)$ on a suitably chosen class of k-chains C_k. It is for this reason we term our method the minimax method.

It is not to be expected that we can treat the minimax principle completely in this paper. We have therefore chosen two of the simplest and most interesting topics. We study first the critical points of a non-degenerate function $f(x_1, \cdots, x_n)$ and secondly the critical extremals of an integral J for the fixed end point problems in the Calculus of Variations.

I

The critical points of functions

1. Hypotheses and definitions. The basic theory underlying the critical point relations can best be illustrated by studying those of a function $f(x_1, \cdots, x_n)$ defined over a region $\mathfrak{S}$ in a euclidean space of points $(x_1, \cdots, x_n)$. We suppose that $f(x)$ is continuous and has continuous first and second derivatives on $\mathfrak{S}$.

A point P on $\mathfrak{S}$ will be called a *critical point* of $f(x)$ if the derivatives f_{x_i} are all zero at the point P. If P is a critical point, then $f(P)$ will be called a *critical value* of $f(x)$. A critical point P will be called *non-degenerate* if the determinant

$$|f_{x_i x_j}| \qquad (i, j = 1, \cdots, n)$$

is different from zero at P. The negative type number of the quadratic form

$$f_{x_i x_j} \pi_i \pi_j \qquad (i, j = 1, \cdots, n)$$

will be called the *type* or the *index* of P as a critical point of $f(x)$. A function f will be said to be *non-degenerate* if all of its critical points are non-degenerate. We shall assume that $f(x)$ is non-degenerate.

We suppose that there is a closed region $\mathfrak{R}$ interior to $\mathfrak{S}$ containing all the critical points of f in its interior. The boundary B of $\mathfrak{R}$ is assumed to be a regular manifold of class C'', that is, one such that for each point P on B there is a neighborhood $\mathfrak{P}$ on B representable in the form

$$\varphi(x_1, \cdots, x_n) = 0,$$

[3] Lefschetz, *Application of chain deformations to critical points and extremals*, Proceedings of the National Academy of Sciences, vol. 21 (1935), pp. 220–221.

[4] See our paper in the Proceedings of the National Academy of Sciences, February 1935, pp. 96–99, with the same title, in which this principle was first formulated. Cf. Birkhoff, *Dynamical Systems*, p. 135; and Morse, p. 272.

where $\varphi(x)$ is continuous and has continuous derivatives of the first two orders and is such that the derivatives φ_{x_i} are not all zero at any point on $\mathfrak{P}$.

On the boundary B of $\mathfrak{R}$ the normal derivative f_N of f along the outer normal is well defined by virtue of the fact that there are no critical points of $f(x)$ on B. We shall suppose that f_N is positive on B. The case in which f_N is not necessarily positive will be considered in §4 below.

It is clear that there are but a finite number of critical points of f in $\mathfrak{R}$ since non-degenerate critical points are necessarily isolated. We may suppose without loss of generality that if P_1, P_2 are two distinct critical points of $f(x)$, then $f(P_1) \neq f(P_2)$. If this were not so, we could alter f slightly in a neighborhood of P_1 so that $f(P_1) \neq f(P_2)$ without changing the type number of P_1 or introducing new critical points. Similarly we can modify f so that in a sufficiently small neighborhood of a critical point P of type k the function f takes the form

$$f = c - x_1^2 - x_2^2 - \cdots - x_K^2 + x_{K+1}^2 + \cdots + x_n^2$$

for a suitable choice of the coördinates (x). The proofs of these facts have been given by Morse and Van Schaack.[5]

The cycles and chains here used are singular cycles and chains taken modulo 2. The number R_k of linearly independent non-bounding k-cycles on $\mathfrak{R}$ in a maximal set is called the k-th connectivity of $\mathfrak{R}$.

2. Critical point relations. In this section we shall give an intuitive proof of the following theorem due to Morse.

THEOREM 2.1. *Let M_k be the number of critical points of type k and R_k the k-th connectivity of the region $\mathfrak{R}$. The following relations hold:*

$$\begin{aligned} M_0 &\geqq R_0, \\ M_1 - M_0 &\geqq R_1 - R_0, \\ M_2 - M_1 + M_0 &\geqq R_2 - R_1 + R_0, \qquad (2.1) \\ &\cdots\cdots\cdots\cdots \\ M_n - M_{n-1} + \cdots + (-1)^n M_0 &= R_n - R_{n-1} + \cdots + (-1)^n R_0. \end{aligned}$$

The first inequality $M_0 \geqq R_0$ in (2.1) is, of course, well known to all students of mathematics. It follows from the fact that $f(x)$ has at least one minimum point in each of its R_0 connected pieces. We can obtain this result in another way which is more complicated but which has the advantage that it can be generalized so as to obtain the remaining relations (2.1). Let C_0 be a 0-cycle on $\mathfrak{R}$ and $F(C_0)$ the maximum value of $f(x)$ on C_0. The inequality $M_0 \geqq R_0$ can be obtained by finding the number of non-equivalent 0-cycles which afford a relative minimum to the functional $F(C_0)$. We generalize this method as follows. Let $F(C_k)$ be the maximum value of $f(x)$ on a k-cycle or a k-chain C_k.

[5] *The critical point theory under general boundary conditions*, Annals of Mathematics, vol. 35 (1934), pp. 547–550.

If the class of admissible k-chains C_k is suitably chosen, the number of non-equivalent k-chains which afford a minimum to $F(C_k)$ is equal to the number M_k of critical points of type k, and the remaining inequalities (2.1) follow readily. This section will be devoted to developing these ideas.

We begin with the 0-cycles C_0 on $\mathfrak{R}$. We admit only such 0-cycles C_0 as are non-bounding on the domain $f \leqq F(C_0)$. Consider now an admissible 0-cycle C_0. If we deform C_0 continuously without increasing $F(C_0)$, we finally obtain a 0-cycle C_0' which cannot be further deformed into a 0-cycle on $f < F(C_0')$ without increasing $F(C_0')$ and which is such that $f(P) = F(C_0')$ at only one point on C_0'. Such a cycle will be called a minimum 0-cycle. The point on C_0' at which $f(P) = F(C_0')$ is clearly a minimum point of $f(x)$. Let us denote by $\mathfrak{N}_0$ the class of all minimum 0-cycles on $\mathfrak{R}$. Each zero cycle in $\mathfrak{N}_0$ determines a unique critical point P of $f(x)$ on $\mathfrak{R}$. But to every minimum point P there will correspond in general infinitely many 0-cycles C_0 having $f(P) = F(C_0)$. These 0-cycles must be considered as equivalent if they are to be used as a count of the minimum points of $f(x)$. We accordingly define two 0-cycles C_0, C_0' in $\mathfrak{N}_0$ to be equivalent if $F(C_0) = F(C_0')$ and their sum $C_0 + C_0'$ is homologous on $f \leqq F(C_0)$ to zero or to the 0-cycles on $f < F(C_0)$. It is clear that for the case here considered two 0-cycles C_0, C_0' in $\mathfrak{N}_0$ are equivalent if and only if $F(C_0) = F(C_0')$. Hence the number N_0 of non-equivalent 0-cycles in $\mathfrak{N}_0$ is equal precisely to the number M_0 of critical points of minimum type. Moreover,

$$M_0 = N_0 \geqq R_0, \tag{2.2}$$

since every non-bounding 0-cycle on $\mathfrak{R}$ can be admissibly deformed into a 0-cycle in $\mathfrak{N}_0$. This proves the first inequality (2.1). For reasons which will appear later it is convenient to refer to the class $\mathfrak{N}_0$ of 0-cycles as the class $\mathfrak{M}_0$ of 0-chains and to denote by M_0 the number of non-equivalent 0-chains in $\mathfrak{M}_0$.

In order to extend the method described above so as to obtain further critical point relations, let us first examine the nature of a critical point P of type 1. We shall suppose that the coördinate system (x) has been chosen so that $(x) = (0)$ at P and that the function $f(x)$ takes the form

$$f = c - x_1^2 + x_2^2 + \cdots + x_n^2$$

in a neighborhood of $\mathfrak{P}$ about P. Consider the 1-chain C_1 defined by the relations

$$x_1^2 \leqq h, \qquad x_2^2 + \cdots + x_n^2 = 0. \tag{2.3}$$

For h sufficiently small the 1-chain C_1 will lie in $\mathfrak{P}$ and moreover the maximum value $F(C_1)$ of $f(x)$ on C_1 is attained only at the critical point P. Now it may be possible to join the end points of the arc C_1 by a second arc C_1' on the domain $f < f(P)$. The closed arc $C_1'' = C_1 + C_1'$ then forms a 1-cycle on $\mathfrak{R}$ which cannot be deformed continuously into a 1-cycle on $f < F(C_1'')$ without increasing $F(C_1'')$. Moreover, $f(Q) = F(C_1'')$ at a point Q on C_1'' only in case the point Q is coincident with the critical point P. Thus we see that the cycle C_1'' can be

considered to be a minimizing 1-cycle for the functional $F(C_1)$. The cycle C_1'' is clearly a linking cycle in the Morse sense.[6]

If the boundary of the 1-chain C_1 defined by (2.3) is not homologous to zero on the domain $f < F(C_1)$, clearly it is homologous on this domain to a 0-cycle C_0 in the class $\mathfrak{N}_0$ of minimizing 0-cycles for $F(C_0)$. Let C_1' be the 1-chain joining the ends of C_1 with the 0-cycle C_0. The 1-chain $C_1'' = C_1 + C_1'$ has properties analogous to those of the 1-cycle C_1'' described in the last paragraph. For example, the 1-chain cannot be deformed continuously into a 1-chain on $f < F(C_1'')$ without increasing $F(C_1'')$ or $F(C_0)$. Moreover, the only point Q on C_1'' at which $f(Q) = F(C_1'')$ is the critical point P under consideration. It should be noted that for the 1-chain C_1'' here constructed the boundary C_0 is not homologous to zero on the domain $f < F(C_1'')$. Here again C_1'' appears as a minimizing 1-chain for the functional $F(C_1)$.

In view of the above remarks it would seem that one should be able to obtain all the critical points of $f(x)$ of type 1 by minimizing $F(C_1)$ on a suitably chosen class of 1-chains. This is indeed the case. We admit, for obvious reasons, only a special class of 1-chains, which we shall term *admissible 1-chains*. A 1-cycle C_1 will be admitted if it is non-bounding on the domain $f \leqq F(C_1)$. A 1-chain possessing a boundary will be admitted if its boundary is in $\mathfrak{N}_0$. In order to find the minimizing cycles for $F(C_1)$, we associate with each admissible 1-chain C_1 a set of deformations, called *admissible deformations*, which never increase $F(C_1)$; nor do they increase $F(C_0)$ if the boundary C_0 of C_1 exists. These deformations are to be continuous, except for the fact that they may subdivide C_1 into a finite number of parts subject to the following restrictions. The image of a 1-cycle C_1 under a deformation must be homologous to C_1 on the domain $f \leqq F(C_1)$. The image C_1' of a 1-chain C_1 having a boundary C_0 must be related to C_1 as follows. Let C_0' be the boundary of C_1' and let C_1'' be a 1-chain on $f \leqq F(C_0)$ bounded by the 0-cycle $C_0 + C_0'$. The 1-cycle $C_1 + C_1' + C_1''$ must be homologous to zero on the domain $f \leqq F(C_1)$ if our deformation is to be admissible.

Consider now an admissible 1-chain C_1. By means of an admissible deformation we can deform C_1 into a 1-chain C_1' which cannot be further deformed into a 1-chain on the domain $f < F(C_1')$ and which has $f(Q) = F(C_1')$ only at the points Q on C_1' which are on a single point P on $\mathfrak{N}$. The point P can be shown to be a critical point of type 1. The class of all 1-chains having the same properties as C_1' will be denoted by $\mathfrak{M}_1$. Each 1-chain C_1 in $\mathfrak{M}_1$ has associated with it a unique critical point P of type 1 having $f(P) = F(C_1)$. However, to each critical point P of type 1 there corresponds in general infinitely many 1-chains C_1 in $\mathfrak{M}_1$ having $F(C_1) = f(P)$. These 1-chains must be regarded as equivalent if they are to be used as a count of the critical points. Hence we shall agree to call two 1-chains C_1, C_1' in $\mathfrak{M}_1$ equivalent if $F(C_1) = F(C_1')$. This definition of equivalence is not sufficiently general to be applicable in the case for which there may be more than one critical point corresponding to each critical value.

[6] Cf. Morse, p. 158.

We shall accordingly say that two 1-cycles C_1, C_1' in $\mathfrak{M}_1$ are *equivalent* if $F(C_1) = F(C_1')$ and their sum $C_1 + C_1'$ is homologous on $f \leqq F(C_1)$ to zero or to the 1-cycles on the domain $f < F(C_1)$. Two 1-chains C_1, C_1' in $\mathfrak{M}_1$ not both 1-cycles will be said to be *equivalent* if $F(C_1) = F(C_1')$, if the boundary of their sum bounds a 1-chain C_1'' on $f < F(C_1)$, and finally if the 1-cycle $C_1 + C_1' + C_1''$ is homologous on $f \leqq F(C_1)$ to zero or to the 1-cycles on $f < F(C_1)$. It will be seen in the next section that our two definitions of equivalence are the same for the case here discussed. It follows that the number of non-equivalent 1-chains in $\mathfrak{M}_1$ in a maximal set is equal precisely to the number M_1 of critical points of type 1.

The number of non-equivalent 1-chains in $\mathfrak{M}_1$ can be evaluated in a second way. To do so let us denote by $\mathfrak{N}_1$ the class of all 1-cycles in $\mathfrak{M}_1$ and the maximum number of non-equivalent 1-cycles in $\mathfrak{N}_1$ by N_1. It is clear that every non-bounding 1-cycle in $\mathfrak{R}$ can be deformed by an admissible deformation into a 1-cycle in $\mathfrak{N}_1$. Hence we have $N_1 \geqq R_1$, where R_1 is the linear connectivity of $\mathfrak{R}$. Moreover, there are exactly $N_0 - R_0$ non-equivalent bounding 0-cycles in $\mathfrak{N}_0$ in a maximal set. Let C_0 be one of these and let b be the greatest lower bound of the values of $F(C_1)$ on the 1-chain C_1 bounded by C_0. Clearly any 1-chain C_1 with $F(C_1)$ near b can be deformed admissibly into one for which $F(C_1) = b$. The resulting chain is in $\mathfrak{M}_1$ and is bounded by C_0. There are accordingly $N_0 - R_0$ non-equivalent 1-chains of this type. Moreover, no linear combination of these 1-chains is equivalent to a 1-cycle in $\mathfrak{N}_1$. It follows readily that there are exactly $N_1 + N_0 - R_0$ non-equivalent chains in $\mathfrak{M}_1$. But the number of non-equivalent chains in $\mathfrak{M}_1$ was seen above to be equal to the number M_1 of critical points of type 1. Hence we have

$$(2.4) \qquad M_1 \geqq N_1 \geqq R_1, \qquad M_1 - N_1 = N_0 - R_0.$$

The second relation (2.1) follows readily from the second relation (2.4) by replacing N_0 by M_0 and N_1 by R_1.

The above arguments can be extended inductively to any dimension. We suppose that the classes $\mathfrak{M}_k$, $\mathfrak{N}_k$ of minimizing k-chains and k-cycles for $F(C_k)$ have been constructed for $k = 0, 1, \cdots, j - 1$. For these values of k the numbers M_k, N_k of non-equivalent k-chains in $\mathfrak{M}_k$, $\mathfrak{N}_k$ respectively satisfy the relations

$$(2.5) \quad M_k \geqq N_k \geqq R_k, \quad M_k - N_k = N_{k-1} - R_{k-1} \quad (k = 1, \cdots, j - 1).$$

The proof that these relations hold for $k = j$ can be established by precisely the same arguments as those made for $k = 1$. We merely need to make the obvious changes such as replacing 1 by j and 0 by $j - 1$ whenever they occur. For this reason we shall give here only the important definitions and ideas used in the development.

In order to establish the relations (2.5) for $k = j$ we minimize $F(C_j)$ in the class of admissible j-chains. A j-cycle C_j will be termed *admissible* if it is non-bounding on the domain $f \leqq F(C_j)$. A j-chain C_j having a boundary C_{j-1} will

be termed *admissible* if C_{j-1} is in $\mathfrak{N}_{j-1}$. An admissible deformation is one which deforms an admissible j-chain C_j without increasing $F(C_j)$, and which deforms the boundary of C_j admissibly. An admissible deformation is to be continuous except that we admit the possibility of C_j breaking up into a finite number of pieces subject to the following restrictions. If C_j is a cycle, its image under an admissible deformation must be homologous to C_j on the domain $f \leqq F(C_j)$. If C_j is a j-chain bounded by a $(j-1)$-cycle C_{j-1} in $\mathfrak{N}_{j-1}$, the image C'_j of C_j and the boundary C'_{j-1} of C'_j must be related to C_j, C_{j-1} as follows. Let C''_j be a j-chain on $f \leqq F(C_{j-1})$ bounded by the $(j-1)$-cycle $C_{j-1} + C'_{j-1}$. The j-cycle $C_j + C'_j + C''_j$ must be homologous to zero on the domain $f \leqq F(C_j)$. We now define a j-chain C_j to be in the class $\mathfrak{M}_j$ if C_j cannot be admissibly deformed into a j-chain on $f < F(C_j)$, and if $f(Q) = F(C_j)$ only at the points Q which are on a single point P on $\mathfrak{R}$. The point P can be seen to be a critical point of type j. Two j-cycles C_j, C'_j in $\mathfrak{M}_j$ will be called *equivalent* if $F(C_j) = F(C'_j)$ and the j-cycle $C_j + C'_j$ is homologous on $f \leqq F(C_j)$ to zero or to the j-cycles on $f < F(C_j)$. Two j-chains C_j, C'_j which are not both j-cycles will be called *equivalent* if $F(C_j) = F(C'_j)$, if the boundary of $C_j + C'_j$ bounds a j-chain C''_j on $f < F(C_j)$, and finally if the j-cycle $C_j + C'_j + C''_j$ is homologous on $f \leqq F(C_j)$ to zero or to the j-cycles on $f < F(C_j)$. For the case here considered two j-chains C_j, C'_j in $\mathfrak{M}_j$ can be seen to be equivalent if and only if $F(C_j) = F(C'_j)$. The number of non-equivalent j-chains in $\mathfrak{M}_j$ in a maximal set is therefore equal precisely to the number M_j of critical points of $f(x)$ of type j. Let $\mathfrak{N}_j$ be the class of all j-cycles in M_j, and N_j be the maximum number of non-equivalent j-cycles in $\mathfrak{N}_j$. Since each non-bounding j-cycle on $\mathfrak{R}$ can be deformed into one in $\mathfrak{N}_j$, we have $N_j \geqq R_j$. Moreover, one can see as in the case $j = 1$ that there are exactly $N_{j-1} - R_{j-1}$ non-equivalent j-chains in $\mathfrak{M}_j$, no linear combination of which is equivalent to a j-cycle in $\mathfrak{N}_j$. The number M_j of non-equivalent j-chains in $\mathfrak{M}_j$ is accordingly equal to $N_j + N_{j-1} - R_{j-1}$. The relations (2.5) accordingly hold for $k = j$. From the relations (2.5) and (2.2) it follows readily that

$$M_k - M_{k-1} + \cdots + (-1)^k M_0 = N_k - R_{k-1} + R_{k-2} - \cdots + (-1)^k R_0,$$

$$M_k - M_{k-1} + \cdots + (-1)^k M_0 \geqq R_k - R_{k-1} + \cdots + (-1)^k R_0,$$

the equality holding if and only if $N_k = R_k$. From the relations (2.5) it is clear that $N_k = R_k$ if either $M_{k+1} = R_{k+1}$ or $M_k = R_k$. In particular $N_n = R_n$, as one readily verifies.

Thus we see that all critical point relations can be obtained by a minimax principle. We minimize the maximum $F(C_k)$ of $f(x)$ on the k-chains of a suitably chosen class. In order to complete the arguments made above we must establish the following statements.

I. *Every k-cycle C_k on $\mathfrak{R}$ which is non-bounding on $f \leqq F(C_k)$ can be deformed admissibly into a k-cycle in $\mathfrak{N}_k$.*

II. *Every k-chain C_k on $\mathfrak{R}$ whose boundary is in $\mathfrak{N}_{k-1}$ can be admissibly deformed into one having an equivalent boundary.*

III. *The points on a k-chain C_k in $\mathfrak{M}_k$ at which $f(P) = F(C_k)$ correspond to a single critical point P of $f(x)$ of type k. Two k-chains C_k, C_k' in $\mathfrak{M}_k$ having $F(C_k) = F(C_k')$ are equivalent. Each critical point P of type k has associated with it at least one k-chain C_k in $\mathfrak{M}_k$ such that P is on C_k and $f(P) = F(C_k)$.*

These statements will be established in the next section.

3. Proofs of the above three statements. The proofs of the three statements made at the end of the last section depend on two lemmas, the first of which is the following

LEMMA 3.1. *Every admissible k-chain C_k on $\mathfrak{R}$ can be admissibly deformed into a k-chain C_k' such that the points on C_k' at which $f = F(C_k')$ are in an arbitrarily small neighborhood $\mathfrak{P}$ of a critical point P and such that the boundaries of C_k and C_k', if they exist, are equivalent. The points of C_k' not in $\mathfrak{P}$ are on the domain $f < f(P)$.*

To prove this theorem we use the orthogonal trajectories to the hypersurfaces $f =$ constant. These trajectories are the solutions of the differential equations

$$dx_i/dt = f_{x_i} \qquad (i = 1, \cdots, n),$$

and are well defined except at the critical points of f. Through any ordinary point of f there passes one and but one of these trajectories.

We now define a deformation Δ_b. To do so, we consider a k-chain C_k. We suppose that there are no critical points P on C_k at which $f(P) = F(C_k)$. Let b be a value of f such that each point P of C_k on the domain $f \geqq b$ can be joined to a unique point P' on $f = b$ by means of an orthogonal trajectory. As the time t varies from 0 to 1, a point P on $f > b$ moves towards P' along the trajectory PP' at a rate equal to its initial length. The points on $f \leqq b$ are held fast. We term this deformation Δ_b. It carries C_k continuously into a k-chain C_k' on the domain $f \leqq b$. The boundary C_{k-1} of C_k, if it exists, is unaltered since b clearly exceeds $F(C_{k-1})$.

A second deformation Δ_b' can now be defined as follows. Let b be a critical value of $f(x)$ corresponding to a critical point P. The orthogonal trajectories through points on $f = b$ at distances ρ, 2ρ from P together with the surfaces $f = b \pm \epsilon$ form tubular neighborhoods T_1, T_2 of P which lie within any prescribed neighborhood $\mathfrak{P}$ of P, provided the positive constants ρ, ϵ are taken sufficiently small. We deform points outside T_2 according to the deformation $\Delta_{b-\epsilon}$. Points inside T_1 are held fast. A point Q in T_2 but not in T_1 is deformed as follows. Let P', Q' be the points of $f = b$, $f = b - \epsilon$, respectively, on the orthogonal trajectory of $f = b$ through Q. The point Q moves towards Q' on this trajectory at a rate equal to $(d - \rho)/\rho$ times the length of the trajectory QQ', where d is the distance from P' to P. The deformation so defined is admissible and will be denoted by Δ_b'.

The deformation described in the lemma can be made by applying successively deformations of the type Δ_b and Δ_b'. The lemma is established.

Our second lemma deals with deformations in a neighborhood of a critical point P of type k. We may suppose that the coördinate system (x) has been

chosen so that P is the point $(x) = (0)$ and that in a neighborhood $\mathfrak{P}$ of P the function $f(x)$ takes the form

$$f = b - x_1^2 - x_2^2 - \cdots - x_k^2 + x_{k+1}^2 + \cdots + x_n^2 . \tag{3.1}$$

Let S_ρ be a spherical neighborhood

$$x_1^2 + \cdots + x_n^2 \leqq \rho^2$$

of P within $\mathfrak{P}$. We are especially interested in the k-chain

$$x_1^2 + \cdots + x_k^2 \leqq \rho^2 , \qquad x_{k+1}^2 + \cdots + x_n^2 = 0 , \tag{3.2}$$

which we shall denote by C_k. Its boundary is denoted by C_{k-1}. A j-chain C'_j in S_ρ with its boundary C'_{j-1} on $f < b$ will be said to be equivalent to C_k if $j = k$ and if C'_{k-1} is homologous to C_{k-1} in S_ρ on $f < b$.

Lemma 3.2. *Let C'_j be a j-chain interior to S_ρ whose boundary C'_{j-1} is on the domain $f < b$. If C'_j is not equivalent to C_k, then C'_j can be admissibly deformed on S_ρ into a j-chain C''_j on $f < b$ having the same boundary. If C'_j is equivalent to C_k, then C_j cannot be so deformed, but can be deformed into a k-chain C''_k on $f \leqq b$ having the same boundary and having $f = b$ only at the points on C_k which are on the critical point P.*

To prove this, choose ϵ so small that $b - 9\epsilon^2 > F(C'_{j-1})$, and let $S_\epsilon, S_{2\epsilon}, S_{3\epsilon}$ be three spherical neighborhoods of P of radius $\epsilon, 2\epsilon, 3\epsilon$ respectively. By means of the deformation Δ'_b we may deform C'_j admissibly so that all points of C'_j which are on the domain $f \geqq b - \epsilon^2$ lie in $S_{2\epsilon}$. We now make the following deformation.[7] The points on C_j outside and on the boundary of the sphere $S_{3\epsilon}$ are held fast. Inside the sphere $S_{3\epsilon}$ the coördinates $x_1, \cdots, x_k$ of a point (x) are held fast. Inside $S_{2\epsilon}$ the coördinates $x_j (j > k)$ move towards zero as the time t varies from zero to 1 at a rate equal to the absolute value of the initial value of x_j. In the region between $S_{2\epsilon}$ and $S_{3\epsilon}$ the coördinates $x_j (j > k)$ move towards zero at a rate equal d/ϵ times the absolute value of the initial value of x_j, where d denotes the distance from the point (x) to the boundary of $S_{3\epsilon}$.

Under the deformation just described the j-chain C'_j is admissibly deformed into a j-chain C''_j on $f \leqq b$ and having the same boundary. The portion of C''_j which is in S_ϵ lies on the k-chain

$$0 \leqq x_1^2 + \cdots + x_k^2 \leqq \epsilon^2 , \qquad x_{k+1}^2 + \cdots + x_n^2 = 0 \tag{3.3}$$

by virtue of our deformation. Suppose now that C'_j, and hence also C''_j, is not equivalent to C_k. If $F(C''_j) < b$, our lemma is established. Hence we need to consider only the case $F(C''_j) = b$. In this case we may suppose that C''_j is divided finely enough so that there is a $(j - 1)$-cycle $\bar{C}_{j-1}$ which is composed of cells of C''_j on the domain

$$0 < x_1^2 + \cdots + x_k^2 \leqq \epsilon^2 , \qquad x_{k+1}^2 + \cdots + x_n^2 = 0 , \tag{3.4}$$

[7] Cf. Morse, p. 170.

and which together with the boundary of C''_j bounds a j-chain γ_j on $f < b$ composed of cells of C''_j. The $(j - 1)$-cycle $\bar{C}_{j-1}$ accordingly bounds the j-chain $\bar{C}_j = C''_j + \gamma_j$ on (3.3). Now the connectivities of the domain (3.4) are those of a $(k - 1)$-sphere, as will be seen below. It follows that $\bar{C}_{j-1}$ bounds a j-chain γ'_j on the domain (3.4), since by construction it is not homologous to the boundary of (3.3). We now think of C''_j as being equal to the sum of the chains $\gamma''_j = \gamma_j + \gamma'_j$ and $\gamma'_j + \bar{C}_j$. The second of these chains is a j-cycle on the domain (3.3) and can be deformed into the null j-cycle in many different ways without increasing $F(C''_j)$. Thus we see that in this case C''_j can be deformed into a j-chain γ''_j on the domain $f < b$. The lemma is therefore true as stated for the case in which C'_j is not equivalent to C_k. If C'_j is equivalent to C'_k, then C''_j can be constructed as before. In this case we must have $F(C''_j) = b$, since otherwise the boundary of C_k would be homologous to zero in S_ρ on the domain $f < b$, which is not the case. For the same reason C''_j cannot be deformed into a j-chain on $f < b$ with the same boundary. The last statement in the lemma is accordingly true.

The proof of the lemma will be complete if we show that the connectivities of the domain (3.4) are those of a $(k - 1)$-sphere. This is immediate, since we can deform the domain (3.4) into the $(k - 1)$-sphere

$$(3.5) \qquad x_1^2 + \cdots + x_k^2 = \epsilon^2, \qquad x_{k+1}^2 + \cdots + x_n^2 = 0$$

as follows. As the time t varies from 0 to 1 let each point P move along the trajectories through P toward the point Q on (3.5) on the same radius at a rate equal to the initial length of the radius PQ.[7] This completes the proof of Lemma 3.2.

Statements I, II, and the first part of III made at the end of the last section follow at once from Lemmas 3.1 and 3.2. The last statement in III follows immediately from the fact that the boundary C_{k-1} of the k-chain C_k given by the relation (3.2) is homologous on the domain $f < f(P)$ to zero or to a cycle C'_{k-1} in $\mathfrak{N}_{k-1}$. Let C'_k be a k-chain on $f < f(P)$ bounded by C_{k-1} or by $C_{k-1} + C'_{k-1}$ as the case may be. The chain $C''_k = C_k + C'_k$ is clearly a k-chain in $\mathfrak{M}_k$ having $f(P) = f(C''_k)$, and the last statement in III is established.

It remains to prove that two k-chains C'_k, C''_k in $\mathfrak{M}_k$ having $F(C'_k) = F(C''_k)$ are equivalent. To do so let P be the critical point of type k on C'_k and C''_k such that $f(P) = F(C'_k) = F(C''_k)$, and suppose that f is in the form (3.1) in a neighborhood $\mathfrak{P}$ of P. We construct the region S_ρ and the k-chain (3.2) as before, and denote the chain by C_k and its boundary by C_{k-1}. If the chain C'_k is finely enough divided, there is a k-chain γ'_k composed of cells of C'_k in S_ρ which is equivalent to C_k in the sense described in the paragraph preceding Lemma 3.2. Otherwise the k-chain C'_k could be admissibly deformed on the domain $f < f(P)$ by Lemma 3.2. Similarly there is a k-chain γ''_k composed of cells of C''_k which is equivalent to C_k. The boundaries γ'_{k-1}, γ''_{k-1} of γ'_k, γ''_k are homologous on $f < f(P)$ in S_ρ to the boundary C_{k-1} of C_k, and hence homol-

ogous to each other on the same domain. Let γ_k be the k-chain bounded by $\gamma'_{k-1} + \gamma''_{k-1}$ in S_ρ on $f < f(P)$. The k-chain $C'_k + C''_k$ can be written in the form

$$C'_k + C''_k = (\gamma'_k + \gamma''_k + \gamma_k) + (C'_k + C''_k + \gamma'_k + \gamma''_k + \gamma_k).$$

The k-chain in the first parentheses is a k-cycle which is homologous to zero on the domain S_ρ. The k-chain in the second parentheses is a k-chain $\bar{C}_k$ on the domain $f < f(P)$, and hence is homologous to the k-cycles on this domain if C'_k, C''_k are both k-cycles. If C'_k, C''_k are not both k-cycles, then $\bar{C}_k$ is a k-chain on $f < F(C'_k)$ bounded by the boundary of the chain $C'_k + C''_k$. Moreover, the k-cycle $C'_k + C''_k + \bar{C}_k$ is clearly homologous to zero on $f \leqq F(C'_k)$. Thus we see that in either case C'_k and C''_k are equivalent. This completes the proof of Theorem 2.1.

4. **General boundary conditions.** In the above treatment we assumed that the normal derivative f_N along the outer normal is positive on the boundary B of $\mathfrak{R}$. This assumption was made only for convenience. If this assumption does not hold, some of the admissible k-chains C_k will, in general, be deformed by an admissible deformation into the region A of B where f_N is negative. If this k-chain C_k is deformed so that $F(C_k)$ is a minimum and so that $f(P) = F(C_k)$ at essentially one point on C_k, this point will not be a critical point of $f(x)$ in the usual sense, but will be a critical point of the function $g(y_1, \cdots, y_{n-1})$ defined by $f(x)$ on the region A. It is clear, therefore, that if the relations (2.1) are to hold in this case we must also include the critical points of $g(y)$ on A as well as those of $f(x)$ on $\mathfrak{R}$. Hence in this case we define M_k to be the number of critical points of type k of $f(x)$ on $\mathfrak{R}$ plus the number of type k of $g(y)$ on A. The remainder of the proof is as before. We assume, of course, that the critical points of $g(y)$ are non-degenerate and that to each critical value of f and g there corresponds but one critical point of f or g. The results here given can also be extended at once to non-singular functions defined on a closed Riemannian manifold. These cases have been discussed by Morse and Van Schaack[8] using a different method.

5. **The degenerate case.** The methods used above can readily be extended to the case in which critical points of $f(x)$ are degenerate. We shall only briefly indicate how this can be done. The only difficulty which arises is that of deforming a k-chain down onto a critical point or a set of critical points. If this can be done, the arguments can be made as above, and we define the count of the critical points of type k to be equal to the number M_k of non-equivalent k-chains in $\mathfrak{M}_k$. The number M_k so obtained will necessarily satisfy the conditions (2.1).

If an admissible k-chain cannot be deformed down onto a set of critical points, we can modify our procedure somewhat and define the k-cycles and k-chains

[8] Loc. cit., footnote 2.

in $\mathfrak{M}_{k\epsilon}$ to be those whose maximum points are in an ϵ-neighborhood of a set of critical points at which $f = b$ but cannot be admissibly deformed into a k-chain on $f < b$. Under a suitable definition of equivalence of cycles and chains in $\mathfrak{M}_{k\epsilon}$ it can be shown by methods analogous to those used above that the number M_k of equivalent k-cycles and k-chains in $\mathfrak{M}_{k\epsilon}$ is independent of the particular choice of ϵ, for ϵ sufficiently small, and satisfies the relations (2.1). In the non-degenerate case the number M_k so defined is equal to the number of critical points of type k.

The treatment here given is in the large. But the same methods can be applied to neighborhoods of sets of critical points. We obtain thereby a characterization of sets of critical points in the small, which is independent of the particular neighborhood used, at least in the most important cases.

II

Generalized minimax principle

The methods used in the last part can be given an abstract formulation which brings out the essential features of the method. The results of this section were published recently by the authors in a somewhat different form.[9]

6. Generalized minimax principle. Consider now a space Ω on which the ordinary concepts of topology, such as k-cycles, k-chains, non-bounding k-cycles, addition, homologies, etc., are well defined (modulo 2). The number R_k of independent k-cycles in a maximal set of such cycles is called the k-th connectivity of Ω. We admit the possibility of R_k being infinite.

Suppose now that we have given a functional $f(P)$ which is well defined for all points P on Ω. We shall denote by $F(C_k)$ the least upper bound of the values $f(P)$ on the chain C_k.

Our critical point relations will be obtained by minimizing the functional $F(C_k)$ on a suitably chosen set of k-chains. A k-cycle C_k will be admitted only in case C_k is not homologous to zero on the domain $f \leqq F(C_k)$. We associate with each admissible k-cycle a set of deformations, called *admissible deformations*, which never increase $F(C_k)$ and which deform C_k into a k-cycle which is homologous to C_k on the domain $f \leqq F(C_k)$. A minimum k-cycle C_k is defined to be one which cannot be deformed admissibly on $f \leqq F(C_k)$ into a k-cycle on the domain $f < F(C_k)$. The class of all minimum k-cycles will be denoted by $\mathfrak{N}_k$. Two k-cycles C_k, C'_k in $\mathfrak{N}_k$ will be called *equivalent* if $F(C_k) = F(C'_k)$ and their sum is homologous on $f \leqq F(C_k)$ to zero or to the k-cycles on the domain $f < F(C_k)$.

A k-chain C_k bounded by $(k - 1)$-cycle C_{k-1} will be admitted if its boundary C_{k-1} is in $\mathfrak{N}_{k-1}$. We associate with such a k-chain C_k a set of deformations, called *admissible deformations*, which never increase $F(C_k)$, which deform its boundary C_{k-1} admissibly, and which deform C_k into an admissible k-chain C'_k

[9] Loc. cit.

related to C_k as follows. Let C'_{k-1} be the boundary of C'_k, and let C''_k be a k-chain on the domain $f \leqq F(C_{k-1})$ bounded by the k-cycle $C_{k-1} + C'_{k-1}$. The k-cycle $C_k + C'_k + C''_k$ must be homologous to zero on $f \leqq F(C_k)$, if our deformation is to be admissible.

By a minimum k-chain C_k will be meant one which cannot be admissibly deformed on the domain $f \leqq F(C_k)$ into a k-chain on $f < F(C_k)$. The class of all minimum k-chains and k-cycles ($k > 0$) will be denoted by $\mathfrak{M}_k$. We set $\mathfrak{M}_0 = \mathfrak{N}_0$. Two k-chains C_k, C'_k, in $\mathfrak{M}_k$ but not both in $\mathfrak{N}_k$, will be called *equivalent* if $F(C_k) = F(C'_k)$, if the boundary of their sum bounds a k-chain C''_k on $f < F(C_k)$, and if the k-cycle $C_k + C'_k + C''_k$ is homologous on $f \leqq F(C_k)$ to zero or to the k-cycles on $f < F(C_k)$.

We make the following assumptions.

I. *Every admissible k-cycle C_k on Ω can be admissibly deformed into a k-cycle in $\mathfrak{N}_k$.*

II. *Every admissible k-chain ($k > 0$) on Ω can be admissibly deformed into a k-chain in $\mathfrak{M}_k$ having the same or an equivalent boundary.*

III. *The value $F(C_k)$ is attained by $f(P)$ on each k-chain C_k in $\mathfrak{M}_k$.*

The following lemma is immediate.

Lemma 6.1. *Let M_k, N_k be respectively the number of non-equivalent k-chains in $\mathfrak{M}_k$, $\mathfrak{N}_k$, and R_k the k-th connectivity of Ω. If the numbers $M_k (k = 0, 1, \cdots, m)$ (m arbitrary) are finite, under assumptions* I, II *the following relations hold.*

$$M_k \geqq N_k \geqq R_k \geqq 0 \qquad (k = 0, 1, \cdots, m), \tag{6.1}$$

$$M_0 = N_0, \qquad M_j - N_j = N_{j-1} - R_{j-1} \qquad (j = 1, \cdots, m), \tag{6.2}$$

$$M_k - M_{k-1} + \cdots + (-1)^k M_0 = N_k - R_{k-1} + R_{k-2} + \cdots + (-1)^k R_0, \tag{6.3}$$

$$M_k - M_{k-1} + \cdots + (-1)^k M_0 \geqq R_k - R_{k-1} + \cdots + (-1)^k R_0, \tag{6.4}$$

the equality holding if and only if $N_k = R_k$. If $M_k = R_k$, or if $M_{k+1} = R_{k+1}$, then $N_k = R_k$. If R_k is infinite, so also are M_k and N_k.

Let C_k be a k-chain in $\mathfrak{M}_k$. Let us deform C_k by an admissible deformation which diminishes f at all points at which it is possible to do so. A point P on the new k-chain C_k at which $f(P) = F(C_k)$ will be called a *critical point of type k*. The count of the critical points of type k is defined to be the number M_k of non-equivalent k-chains in $\mathfrak{M}_k$. We have the following theorem.

Generalized Minimax Principle. *Under hypotheses* I, II, *and* III *the critical points of type k exist and their counts M_k satisfy the relations* (6.4).

The minimax principle here given involves an ideal set. It can be realized at least in the non-degenerate cases considered in this paper. It is not clear that it can be realized in the general degenerate cases. The chief difficulty which arises is the construction of the classes $\mathfrak{M}_k$ and $\mathfrak{N}_k$. However, these classes can in general be approximated by classes $\mathfrak{M}_{ke}$, $\mathfrak{N}_{ke}$, as suggested in §5 above.

III

The fixed end point problem in the Calculus of Variations

One of the numerous applications of the minimax principle described in the last section is to the fixed end point problem in the Calculus of Variations. In this case we are interested in the existence and the classification of the extremals of an integral of the form

$$J = \int_{t_1}^{t_2} f(x, \dot{x})\, dt$$

which join two fixed points A_1 and A_2 on a Riemannian manifold $\mathfrak{R}$. We shall discuss only the non-degenerate case.

7. Hypotheses and definitions. We assume that $f(x, \dot{x})$ is a positive analytic function for all points $(x, \dot{x})$ with (x) on $\mathfrak{R}$ and $(\dot{x}) \neq (0)$. We assume further that for these values of $(x, \dot{x})$

$$f(x, k\dot{x}) = kf(x, \dot{x}) \qquad (k > 0),$$

and that $f(x, \dot{x})$ is positively regular, that is,

$$f_{\dot{x}^i\dot{x}^k}\pi_i\pi_k > 0 \qquad (i = 1, \cdots, n)$$

for all $(\pi) \neq (\rho\dot{x})$.

An arc[10] of class D' which joins the two fixed points A_1 and A_2 will be called an *admissible* arc. The totality of admissible arcs will be defined as our space Ω. The arguments here given hold equally well in case the space Ω is taken to be the totality of admissible arcs for which $J \leqq b$, where b is an arbitrary fixed constant.

We introduce a Fréchet distance $d(E_1, E_2)$ between the arcs E_1 and E_2 on Ω as follows. The geodesic distance beween points P_1 and P_2 on $\mathfrak{R}$ is defined to be the greatest lower bound of the lengths of the arcs on $\mathfrak{R}$ joining P_1 and P_2. Let H be a homeomorphism between the arcs E_1 and E_2 preserving sense and let $\delta(H)$ be the maximum geodesic distance between corresponding points under H. We now define $d(E_1, E_2)$ to be the greatest lower bound of $\delta(H)$ for all sense-preserving homeomorphisms H between E_1 and E_2. Cycles and chains on Ω are to be defined in the manner described by Morse.[11]

8. Critical extremals. An *extremal* is a solution of class C'' of the Euler equations

$$f_{x^i} - (d/dt)f_{\dot{x}^i} = 0 \qquad (i = 1, \cdots, n).$$

[10] An arc $x^i = x^i(t)$ $(t_1 \leqq t \leqq t_0)$ will be said to be of class D' if it is continuous and is composed of a finite number of sub-arcs on each of which $x^i(t)$ have continuous derivatives $\dot{x}^i$, not all zero, for a suitable choice of the parameter t.

[11] Morse, pp. 193–5.

A *critical extremal* is defined to be one which joins the two given points A_1 and A_2. The value of J along a critical extremal will be called a *critical value* of J. In the analytic case here considered there are at most a finite number of critical values of J less than a given constant b. A proof of this fact has been given by Morse.[12]

A critical extremal E will be said to be *non-degenerate* if its end points are not conjugate. We shall assume that the critical extremals are all non-degenerate. In this case critical extremals are isolated and there are but a finite number of extremals corresponding to each critical value.[13]

The *type number* of a non-degenerate critical extremal E is defined to be the sum of the orders of the conjugate points on E of the initial point A. We shall prove the following

THEOREM 8.1. *Let M_k be the number of critical extremals of type k and R_k the k-th connectivity of Ω. For the values M_k, R_k that are finite the relations*

$$M_k \geqq R_k,$$

$$M_k - M_{k-1} + \cdots + (-1)^k M_0 \geqq R_k - R_{k-1} + \cdots + (-1)^k R_0$$

are true. The equality in the last expression holds in case either $M_k = R_k$ or $M_{k+1} = R_{k+1}$. If R_k is infinite, M_k is infinite.

This theorem will follow at once from the generalized minimax principle of §6 if we establish the following facts. The classes $\mathfrak{M}_k$, $\mathfrak{N}_k$ here used are defined as in §6. We denote the maximum value of J on a chain C_k by $J(C_k)$.

I. *Every k-cycle C_k which is non-bounding on the domain $J \leqq J(C_k)$ can be admissibly deformed into a k-cycle in $\mathfrak{N}_k$.*

II. *Every k-chain C_k whose boundary is in $\mathfrak{N}_{k-1}$ can be admissibly deformed into a k-chain in $\mathfrak{M}_k$ having the same or an equivalent boundary.*

III. *The number of non-equivalent k-chains C_k in $\mathfrak{M}_k$ having $J(C_k) = b$ is equal to the number of critical extremals E of type k having $J(E) = b$.*

9. The space Ω_m.

In order to prove the three statements made at the end of the last section it is convenient to study first particular sub-spaces of Ω which we denote by Ω_m and which we shall now define.

We term the value of J along a curve E the *J-length* of E. Let ρ be a constant so small that every extremal of J-length $\leqq 2\rho$ affords a proper minimum to J in the class of all admissible arcs joining its end points. An extremal segment of J-length $\leqq \rho$ will be called an *elementary extremal*. This terminology is due to Morse.

The space Ω_m is now defined as the totality of curves in Ω composed of at most $m + 1$ elementary extremals. The end points of the successive elementary extremals on an arc E in Ω_m form a sequence

$$P_0 = A_1, P_1, \cdots, P_m, P_{m+1} = A_2, \tag{9.1}$$

[12] Loc. cit., p. 199.
[13] Morse, p. 230.

which we call the *vertices* of E. We admit the possibility of successive vertices being coincident.

Lemma 9.1. *Let C_k be any k-chain on Ω. If m is sufficiently large, then C_k can be admissibly deformed into a k-chain on Ω_m.*

To prove this, divide the interval $0 \leqq t \leqq 1$ of the parameter t of the arcs on C_k into $m + 1$ equal segments $t_i t_{i+1}$. For m sufficiently large, the points P_i, P_i' on any arc E of C_k determined by values t_i, t_i' (t_i' on $t_i t_{i+1}$) can be joined by an elementary extremal E_i. Assign to a point P on E_i the parameter value t which divides $t_i t_{i+1}$ in the same ratio as P divides E_i. Let t_i' vary continuously from t_i to t_{i+1}, the arc $P_i P_i'$ on E being replaced by the corresponding extremal E_i. The chain C_k is then deformed admissibly into a k-chain on Ω_m, as one readily verifies. The deformation here used is due to Morse (p. 205).

Thus we see that we can restrict ourselves for the most part to the study of k-cycles and k-chains on Ω_m.

The following lemma is useful in establishing the existence of extremals.

Lemma 9.2. *The space Ω_m is compact.*

For let $\{E_n\}$ be a sequence of curves in Ω_m, and let

$$P_0^{(n)} = A_1, P_1^{(n)}, \cdots, P_m^{(n)}, P_{m+1}^{(n)} = A_2 \tag{9.2}$$

be a set of vertices on the arcs E_n. This sequence of vertices has at least one accumulation set (9.1) such that the points P_i, P_{i+1} can be joined by an elementary extremal. The vertices (9.1) determine an arc E in Ω_m, which is clearly an accumulation curve of the sequence $\{E_n\}$. The lemma is therefore true.

10. The deformation Δ. Our principal deformation which we shall denote by Δ can be defined as follows. Let E be an arc in Ω_m with vertices (9.1). As the time t varies from 0 to 1/2, points $Q_i (i = 0, 1, \cdots, m)$ move on E from the points P_i towards P_{i+1} at a J-rate[14] equal to the J-length of $P_i P_{i+1}$. The vertices

$$A_1, Q_0, \cdots, Q_m, A_2$$

determine a curve in Ω_{m+1} which varies continuously from the curve E to a curve $\bar{E}$ as t varies from 0 to 1/2. As t varies from 1/2 to 1 let the points $P_i' (i = 1, \cdots, m)$ move on $\bar{E}$ from Q_i towards Q_{i-1} at a J-rate equal to the J-length of $Q_{i-1} Q_i$, the point P_0' moving at a J-rate[14] equal to twice the J-length of $Q_0 A_1$. The vertices

$$A_1, P_0', \cdots, P_m', A_2$$

determine a curve which varies continuously on Ω_{m+1} from $\bar{E}$ to a curve E' as t varies from 1/2 to 1. The final curve E' is in Ω_m, since here $P_0' = A_1$. The deformation thus defined will be called $\Delta(t) (0 \leqq t \leqq 1)$. Deformations of this type have been used by Birkhoff and Morse.

[14] The J-rate of P_i is defined as follows. The J-length of the arc $P_{i-1} P_i$ is a function $h(t)$ of the time t. If $h(t)$ is differentiable, then the quantity $h'(t)$ will be called the J-rate of the point P_i. See Morse, p. 199.

The following lemma is immediate.

LEMMA 10.1. *Under the deformation $\Delta(t)(0 \leqq t \leqq 1)$ a curve E on Ω_m is deformed continuously through curves of Ω_{m+1} into a curve E' on Ω_m such that $J(E) \geqq J(E')$, the equality holding if and only if $E \equiv E'$, that is, if and only if E is an extremal. Moreover, k-chains are deformed admissibly under Δ.*

We have the further result

LEMMA 10.2. *Let $\{E_n\}$ be a sequence of curves in Ω_m having a unique limit curve E. Let E'_n, E' be the images of E_n, E under $\Delta(t)$. The curve E' is the unique limit curve of the sequence $\{E'_n\}$.*

To prove this let (9.2) be a set of vertices for the curves E_n. We may assume that these points have been chosen so as to have a unique limit set (9.1), the vertices of E. Let $Q_i^{(n)}$ be the J-mid-points of the arcs P_iP_{i+1}, and $P_i'^{(n)}$ be the J-mid-points of $Q_{i-1}^{(n)}Q_i^{(n)}$. The points

$$P_0'^{(n)} = A_1, P_1'^{(n)}, \cdots, P_m'^{(n)}, P_{m+1}'^{(n)} = A_2$$

are the vertices of the curves E'_n. It is clear that this set has a unique limit set

$$P'_0 = A_1, P'_1, \cdots, P'_m, P'_{m+1} = A_2, \tag{10.1}$$

namely, the J-mid-points of the arcs $Q_{i-1}Q_i$, where Q_i is the unique limit point of $Q_i^{(n)}$. The set (10.1) forms a set of vertices for E' and the lemma is established.

LEMMA 10.3. *Let S be a set of arcs on Ω_m such that the closure of S contains no extremal arc. There exists a positive constant d such that if E is a curve of S and E' its image under $\Delta(t)$, then*

$$J(E) \geqq J(E') + d.$$

For suppose the lemma were false. Then for every positive constant d_n there would exist a curve E_n such that its image E'_n under Δ would satisfy the relation

$$J(E_n) \leqq J(E'_n) + d_n \qquad (n = 1, 2, \cdots). \tag{10.2}$$

The sequence $\{E_n\}$ could be modified so as to have a limit curve E. The sequence $\{E'_n\}$ would then have a unique limit curve E', the image of E under Δ, by Lemma 10.2. From the relation (10.2) we could conclude that $J(E) \leqq J(E')$, and hence that $J(E) = J(E')$, by Lemma 10.1. But this could be true only in case E is an extremal, contrary to our assumption that the closure of S contains no extremal arc. This proves Lemma 10.3.

The following lemma establishes the existence of critical extremals.

LEMMA 10.4. *Let $\{E_n\}$ be a sequence of curves on Ω_m such that E_n is the image of E_{n-1} under $\Delta(t)$. The sequence $\{E_n\}$ has a unique limit curve which is an extremal.*

For by Lemma 10.1 we have

$$J(E_{n-1}) \geqq J(E_n) \geqq 0.$$

It follows that the numbers $J(E_n)$ have a greatest lower bound J_0. Clearly $J(E) = J_0$ for every accumulation curve E of $\{E_n\}$. Moreover, E is an extremal, since its image E' under $\Delta(t)$ is also an accumulation curve of $\{E_n\}$ by Lemma 10.2 and our choice of E_n as the image of E_{n-1}. The uniqueness follows readily with the help of Lemma 10.3 since our critical extremals, being non-degenerate, are isolated.

11. Deformations in a neighborhood of an extremal. Consider now an extremal E of type k. Choose the integer m and if necessary diminish the constant ρ of §9 so that

$$m\rho < J(E) < (m + 1)\rho.$$

Let η be a positive constant so small that every arc E' in Ω_m at a distance $d \leqq \eta$ from E is such that

$$m\rho < J(E') < (m + 1)\rho.$$

Divide E into $m + 1$ sub-arcs of equal J-length by points $P_1, \cdots, P_m$. Through each of these points pass regular analytic manifolds $\pi_1, \cdots, \pi_m$ cutting E orthogonally. Let us denote by Π the class of arcs of Ω_m which lie in the η-neighborhood just described and which have their vertices on the manifolds $\pi_1, \cdots, \pi_m$. We have the following

LEMMA 11.1. *Let C_k be a k-chain on Ω_m such that the arcs on C_k at a distance d, $0 < \delta \leqq d \leqq \eta$, from E are on the domain $J < J(E)$. If the constant δ is sufficiently small, the chain C_k can be admissibly deformed into a k-chain C_k' on Ω such that the arcs on C_k' at a distance $d \leqq \delta$ are on the domain* Π, *the arcs at a distance d ($\delta \leqq d \leqq \eta$) are on $J < J(E)$, and the arcs at a distance $d \geqq \eta$ are identical with those on C_k.*

This deformation can be accomplished as follows. Let $P_1, \cdots, P_m$ be, respectively, the points in which an arc E' in the η-neighborhood of E intersects the surfaces $\pi_1, \cdots, \pi_m$. Let $P_0 = A_1$ and $P_{m+1} = A_2$. If E' is at a distance $d \leqq \eta/2$ from E, we deform E as follows. As the time t varies from 0 to 1, the points P_i' move on E' from P_{i-1} to P_i at a J-rate equal to the J-length of the sub-arc $P_{i-1}P_i$ on E', the arc $P_{i-1}P_i'$ being replaced by the elementary extremal $P_{i-1}P_i'$. If E' is at a distance d, $\eta/2 \leqq d \leqq \eta$, the points P_i' move from P_{i-1} towards P_i at a J-rate equal to $2(\eta - d)/\eta$ times the J-length of $P_{i-1}P_i$. The arcs on C_k at a distance $d \geqq \eta$ from E are held fast. Under this deformation C_k is deformed into a k-chain C_k' having the properties described in the lemma.

Consider now the manifolds $\pi_1, \cdots, \pi_m$ described at the beginning of this section. The point P_i on the manifold π_i is determined by a set of $n - 1$ parameters $v_{i1}, \cdots, v_{i,n-1}$. Let $q = m(n - 1)$ and let the first $n - 1$ variables of the set $u_1, \cdots, u_q$ be the parameters on π_1, the next $n - 1$ the parameters on π_2, and so on. We have further

LEMMA 11.2. *On the domain* Π *the functional J is an analytic function $f(u)$*

of the parameters $u_1, \cdots, u_q$. *The function* $f(u)$ *has a non-degenerate critical point of type* k *at the point* (u_0) *corresponding to the extremal* E.

The first statement is immediate. The second statement can be readily established by elementary means with the help of the positive regularity of our integral J. An explicit proof has been given by Morse.[15]

Let E be a critical extremal of type k. By the last lemma we see that in a sufficiently small neighborhood S of E the integral J is an ordinary function $f(u)$ of q variables. It can readily be seen in a manner analogous to that given in the paragraph preceding Lemma 3.2 that there is a k-chain C_k on S having its boundary C_{k-1} on the domain $J < J(E)$ and having $J = J(C_k)$ only at the arc E on C_k. The cycle C_{k-1} is non-bounding in S on $J < J(E)$. Moreover, every $(k - 1)$-cycle on this domain is homologous to zero or to C_{k-1} on this domain. A j-chain C'_j on S whose boundary is on $J < J(E)$ will be said to be equivalent to C_k if $j = k$ and if the boundary of C'_j is homologous to the boundary of C_k on the domains S and $J < J(E)$.

Lemma 11.3. *A j-chain C'_j on S whose boundary is on the domain $J < J(E)$ and which is not equivalent to C_k can be admissibly deformed into a j-chain on the domain $J < J(E)$ having the same boundary. A j-chain C'_j on S which is equivalent to C_k and has its boundary on $J < J(E)$ cannot be so deformed but can be deformed into one having the same boundary and having $J = J(C'_j)$ only at the arcs which coincide with E.*

This result can be established by an argument like that given in the proof of Lemma 3.2 with the help of Lemma 7.1 of Morse, p. 169.

12. **Proofs of three statements.** The statements I, II, III made at the end of §8 can now be established as follows. Consider any admissible j-chain C_j. By successive application of the deformation Δ of §10 we may deform C'_j so that the arcs on which $J \geqq b$ lie in an arbitrarily small neighborhood of a set ω of critical extremals on which $J = b$. If this neighborhood is sufficiently small, it will consist of a finite number of non-overlapping neighborhoods each of which contains but one extremal and is such that J is a function $f(u)$ as described in Lemma 11.2. Let S be any one of these neighborhoods and let E be the extremal interior to S. Denote the type number of E by k. By successive applications of the deformation Δ the j-chain C_j can be deformed so that the only arcs on C_j at which $J \geqq b$ are those lying in an arbitrarily small closed neighborhood S' of E interior to S. If C_j is finely enough divided, there will be a j-chain C'_j composed of all the cells of C_j interior to and on the boundary of S and having its boundary on the domain between S and S'. According to Lemma 11.3 the j-chain C'_j can be deformed into one on the domain $J < J(E)$ having the same boundary except in the case in which C'_j is equivalent to the chain C_k described in the paragraph preceding Lemma 11.3. If C'_j is equivalent to C_k, then C'_j can be deformed into a chain C''_j containing E, having $J(C''_j) = J(E)$, and having the same boundary as C'_j. Thus we see that the whole chain

[15] Pp. 196–7.

C_j can be deformed admissibly into a j-chain in the class $\mathfrak{M}_j$ such that the points at which $J = J(C_j)$ are extremals of type j. This proves statements I and II.

In order to prove statement III, we first note that each extremal E of type j has associated with it a j-chain C_j in $\mathfrak{M}_j$ such that E is on C_j and $J(E) = J(C_j)$. For, according to the remarks preceding Lemma 11.3, there is one and but one independent $(j - 1)$-cycle C_{j-1} in a sufficiently small neighborhood S of E which is on the domain $J < J(E)$ in S, is non-bounding in this domain, and bounds a j-chain C_j in S containing the extremal E and having $J = J(C_j)$ only at E. The $(j - 1)$-cycle C_{j-1} is homologous on the domain $J < J(E)$ to zero or to a $(j - 1)$-cycle C'_{j-1} in $\mathfrak{N}_{j-1}$ on this domain. Let C''_j be a chain bounded by C_{j-1} or by $C_{j-1} + C'_{j-1}$ as the case may be. The j-chain $C_j = C'_j + C''_j$ is in $\mathfrak{M}_j$, contains E, and has $J = J(C_j)$ only at E.

Consider now a set ω of critical extremals on which $J = b$. Let $E_1, \cdots, E_m$ be the extremals of type j in ω and let $C^1_j, \cdots, C^m_j$ be j-chains of the type described in the last paragraph. These j-chains are clearly non-equivalent. We shall now prove that any j-chain C_j in $\mathfrak{M}_j$ with $J(C_j) = b$ is equivalent to some linear combination of these m j-chains. To do so, deform C_j admissibly, if necessary, so that the set of points on C_j at which $J = J(C_j)$ is composed of some sub-set of the extremals $E_1, \cdots, E_m$, say the extremals $E_1, \cdots, E_r$. Let $S_1, \cdots, S_r$ be neighborhoods of these points chosen as in the last paragraph. Let η be a positive constant so small that the η-neighborhood of E_α lies in S_α $(\alpha = 1, \cdots, r)$. We may suppose that C_j has been so deformed that the points on C_j outside these η-neighborhoods are on the domain $J < b$ and so that C_j cannot be admissibly deformed out of any one of these neighborhoods. If C_j is sufficiently finely divided, there will be a j-chain γ^α_j composed of all the cells of C_j having points in the η-neighborhood in E_α and having its boundary γ^α_{j-1} in S_α on $J < b$. Moreover, the chain C^α_j, if finely enough divided, has a similar j-chain $\bar{C}^\alpha_j$ with boundaries $\bar{C}^\alpha_{j-1}$ on S_α and $J < b$. The $(j - 1)$-cycles γ^α_{j-1}, $\bar{C}^\alpha_{j-1}$ are homologous on the domain $J < b$ in S and hence bound a j-chain β^α_j in S on $J < b$. The chain $\bar{C}_j + \gamma^\alpha_j + \beta^\alpha_j$ forms a j-cycle δ^α_j which is homologous to zero on the domain $J \leqq b$ in S_α. Let

$$C'_j = C^1_j + \cdots + C^r_j, \qquad \delta_j = \delta^1_j + \cdots + \delta^r_j,$$

and consider the relation

$$\delta_j = (C_j + C'_j) + (C_j + C'_j + \delta_j).$$

The j-chain in the last parentheses is clearly on the domain $J < b$. Moreover, by construction δ_j is homologous to zero on the domain $J \leqq b$. Hence from the definition of equivalence in §6 we see that C_j is equivalent to C'_j, as was to be proved. Statement III of §8 is accordingly established.

HARVARD UNIVERSITY.

ON THE FUNDAMENTAL NUMBER OF A RATIONAL GENERALIZED QUATERNION ALGEBRA

By Claiborne G. Latimer

1. Introduction. Let $\mathfrak{A}$ be a rational generalized quaternion algebra, hereafter referred to merely as an algebra. $\mathfrak{A}$ has a basis $1, i, j, ij$:

$$i^2 = -\alpha, \qquad j^2 = -\beta, \qquad ij = -ji,$$

where α, β are integers, neither divisible by the square of a prime. Such a basis will be said to be a normal basis associated with α and β.

Brandt defined the fundamental number d of $\mathfrak{A}$, employing an arbitrarily chosen maximal realm of integrity $\mathfrak{G}$ in his definition, and showed that d is independent of the particular $\mathfrak{G}$ in $\mathfrak{A}$ which is employed, and that two algebras with the same d are equivalent.[1] We shall determine d explicitly in terms of α and β. This gives a simple criterion for the equivalence of two algebras.

Starting with a normal basis, as above, Albert[2] showed by a series of transformations that $\mathfrak{A}$ has such a basis associated with certain integers τ and σ which have the following properties:

(a) τ is a positive prime, $\tau \equiv 3 \pmod 4$;

(b) σ is an integer prime to τ, containing no square factor > 1, and $-\sigma$ is a quadratic residue of τ;

(c) $-\tau$ is a quadratic non-residue of every odd prime factor of σ;

(d) if σ is even, $\tau \equiv 3 \pmod 8$.

From the method by which such a basis is obtained, there is no obvious relation between the initial α, β and the final τ, σ. τ is any one of the infinitude of primes represented by a certain quadratic form, with a finite number of exceptions. σ was not shown to be unique, but if $\mathfrak{A}$ is not a division algebra, it was shown that $\sigma = -1$.

We shall show that $\sigma = d$, and hence is uniquely determined by $\mathfrak{A}$. Also, that τ may be an arbitrarily chosen prime satisfying the four conditions above. We may take τ as the least such prime and thus have a normal basis associated with a pair of integers which are uniquely determined by $\mathfrak{A}$.

2. The determination of d. According to Dickson's[3] definition, a set of integral elements in $\mathfrak{A}$ is a set having certain properties R, C, U, M. It may be

Received May 15, 1935.

[1] *Idealtheorie in Quaternionentheorie*, Mathematische Annalen, vol. 99 (1928), pp. 9, 12.

[2] *Integral domains in rational generalized quaternion algebras*, Bulletin of the American Mathematical Society, vol. 40 (1934), pp. 164–76. In particular, see Theorems 2, 3. In this paper, we replace Albert's τ, σ by $-\tau$, $-\sigma$, respectively.

[3] *Algebras and their Arithmetics*, pp. 141–2.

shown that this definition is equivalent to Brandt's definition of a maximal realm of integrity. Let $\mathfrak{G}$ be such a set, or realm, with a basis $\omega_1, \cdots, \omega_4$. The norm of the general element in $\mathfrak{G}$ is $N(\Sigma x_i\omega_i) = \frac{1}{2}\Sigma g_{ij}x_ix_j = G(x_1, \cdots, x_4)$, where the coefficients of G are rational integers and $g_{ij} = g_{ji}$. By Brandt's definition of d, the determinant $|g_{ij}| = d^2$, and d is positive or negative according as G is definite or indefinite.

Let $\mathfrak{A}$ have a normal basis associated with α and β as in §1. The norm of the general element in $\mathfrak{A}$ is

$$N(y_1 + y_2i + y_3j + y_4ij) = y_1^2 + \alpha y_2^2 + \beta y_3^2 + \alpha\beta y_4^2 \equiv f.$$

If the ω's are expressed as linear functions of 1, i, j, ij, let the matrix of the coefficients be (t_{ij}). Then G is obtained from f by a transformation of determinant $|t_{ij}|$. Hence $d^2 = 16\alpha^2\beta^2|t_{ij}|^2$, and to determine d^2 it is sufficient to find the value of $|t_{ij}|$ for a basis of some $\mathfrak{G}$. It will be observed that f, and hence G, is definite if and only if $\alpha > 0, \beta > 0$. Hence $d > 0$ if and only if $\alpha > 0, \beta > 0$.

Let $\alpha = \alpha_1\delta$, $\beta = \beta_1\delta$, where δ is the positive g.c.d. of α and β. Let $\pm\alpha'$, $\pm\beta'$, $\pm\delta'$ be the largest positive divisors of α_1, β_1, δ, respectively, such that each of the congruences $x^2 + \beta \equiv 0 \pmod{\alpha'}$, $y^2 + \alpha \equiv 0 \pmod{\beta'}$, $z^2 + \alpha_1\beta_1 \equiv 0 \pmod{\delta'}$ has a solution, the signs being so chosen that $A \equiv \alpha_1/\alpha'$, $B \equiv \beta_1/\beta'$, $\Delta \equiv \delta/\delta'$ are all positive.

Suppose $\alpha \equiv \beta \equiv 1 \pmod 2$. For this case, the writer determined a system of basal elements for each of the sets $\mathfrak{G}$ which contain i, j.[4] An examination of these basal elements shows, for each of the subcases treated, that the determinant $|t_{ij}| = (4\alpha'\beta'\delta'\delta)^{-1}$, but when $\alpha \equiv \beta \equiv 1 \pmod 4$, $|t_{ij}| = (2\alpha'\beta'\delta'\delta)^{-1}$. Hence from the above expression for d^2, $d = \pm AB\Delta$, or $d = \pm 2AB\Delta$, according as α and β are not or are both $\equiv 1 \pmod 4$.

Suppose $\alpha \equiv \beta \equiv 0 \pmod 2$. For this case, Darkow[5] found a system of basal elements for each of the sets containing i, j. She treated four subcases, which were designated as types A, B, C, D. From an examination of these basal elements, we find, as in the case where α and β are odd, that $d = \pm 2AB\Delta$ or $d = \pm AB\Delta$ according as $\mathfrak{A}$ is one of the first two or last two types. Setting $\delta_1 = \delta/2$, $\mathfrak{A}$ is one of the last two types if and only if the congruence

$$2\alpha_1\beta_1x^2 + \beta_1\delta_1y^2 + \alpha_1\delta_1z^2 + 2w^2 \equiv 0 \pmod 8$$

[4] *Arithmetics of generalized quaternion algebras*, American Journal of Mathematics, vol. 48 (1926), pp. 57–66. In particular, see pp. 61–2. The notations of the present paper, when α and β are odd, are the same as in the paper cited, except that α, α_1, α', I, J, K of the former paper are here replaced by $-\alpha$, $-\alpha_1$, $-\alpha'$, i, j, ij, respectively.

[5] *Determination of a basis for the integral elements of certain generalized quaternion algebras*, Annals of Mathematics, (2), vol. 28 (1927–8), pp. 263–70. When our α and β are even, most of the notations of the present paper are different from Darkow's. Her α, μ, ν, 2δ, μ', ν', $2\delta'$, M, N, D, e_1, e_2, e_3 are the same as the present $-\alpha$, $-\alpha_1$, β_1, δ, $-\alpha'$, β', δ', A, B, Δ, i, j, ij, respectively.

has a solution with y, z odd.[6] Multiplying the left member by δ_1, we obtain the equivalent congruence

$$\alpha_1\beta_1\delta x^2 + \beta_1 y^2 + \alpha_1 z^2 + \delta w^2 \equiv 0 \pmod 8 \tag{1}$$

which is symmetric in α_1, β_1, δ.

Suppose $\alpha \equiv \beta + 1 \pmod 2$. In the basis $1, i, j, ij$ we may replace i, j by $i_1 \equiv i$, $j_1 \equiv ij/\delta$, respectively, or by $i_2 \equiv ij/\delta$, $j_2 \equiv j$, respectively. The first of these replacements is equivalent to interchanging α_1 and δ; the second, to interchanging β_1 and δ. Since α_1 or β_1 is even, this case may thus be reduced immediately to the preceding case by a proper interchange. We have then the

THEOREM. *Let $\mathfrak{A}$ be a rational generalized quaternion algebra with a basis* $1, i, j, ij$,

$$i^2 = -\alpha, \qquad j^2 = -\beta, \qquad ij = -ji,$$

α and β being integers, neither divisible by the square of a prime. Let $\alpha = \alpha_1\delta$, $\beta = \beta_1\delta$, where δ is the positive g.c.d. of α and β. Let A, B, Δ be the odd positive divisors of α_1, β_1, δ, respectively, as defined above. Then the fundamental number of $\mathfrak{A}$ is $d = \pm AB\Delta$ or $d = \pm 2AB\Delta$. d is positive if and only if $\alpha > 0$, $\beta > 0$. If α and β are odd, d is even if and only if $\alpha_1 \equiv \beta_1 \equiv \delta \pmod 4$. If α or β is even, d is odd if and only if (1) *has a solution in integers with two of y, z, w odd.*

We have the

COROLLARY. *$\mathfrak{A}$ has a normal basis associated with d and τ, where τ is any prime satisfying the conditions* (a) *to* (d) *of* §1 *with $\sigma = d$.*

For by the theorem, the algebra with a normal basis associated with d and τ has the fundamental number d and hence is equivalent to $\mathfrak{A}$.

Brandt stated that an integer is the fundamental number of a division algebra if and only if it may be written in the form $d = (-1)^{n+1}p_1p_2 \cdots p_n$ where the p's are distinct primes. By our theorem, a fundamental number has no square factor > 1. Brandt's statement may then be proved by showing that there is a prime satisfying the conditions (a) to (d), with $\sigma = d$, if and only if d may be written as above.[7]

UNIVERSITY OF KENTUCKY.

[6] Darkow, loc. cit., p. 266, and equation (8), p. 264.

[7] An examination of (1) and of the criteria of the theorem for the parity of d, for the various cases which may arise, will show that these criteria may be replaced by the following: d is odd if and only if $(\alpha_1+\beta_1)(\beta_1+\delta)(\delta+\alpha_1)(\alpha_1+\beta_1+\delta) \equiv 0 \pmod{16}$. (Added in proof.)

ON THE CHARACTERISTIC EXPONENTS IN CERTAIN TYPES OF PROBLEMS OF MECHANICS

By H. E. Buchanan and W. L. Duren, Jr.

1. **Introduction.** Some years ago H. E. Buchanan[1] published a discussion of periodic orbits near the straight line and equilateral triangle positions in the problem of three finite bodies. More recently he has discussed[2] small oscillations of the so-called neutral helium atom near the straight line and equilateral triangle positions. In all four of these problems the characteristic exponents $0, 0, \pm i\omega$, where ω is the angular velocity, occurred. This paper is an attempt to find out whether one could have predicted the appearance of these exponents from the known integrals of the equations.

All of the problems mentioned above were set up in axes rotating uniformly with angular speed ω. The differential equations in each case may be written in the form

$$\begin{aligned}
&\frac{d^2\xi_i}{dt^2} - 2\,\omega\,\frac{d\eta_i}{dt} = \omega^2\,\xi_i + \frac{1}{m_i}\,\frac{\partial U}{\partial \xi_i}\,,\\
(1)\qquad &\frac{d^2\eta_i}{dt^2} + 2\,\omega\,\frac{d\xi_i}{dt} = \omega^2\,\eta_i + \frac{1}{m_i}\,\frac{\partial U}{\partial \eta_i}\,,\\
&\frac{d^2\zeta_i}{dt^2} = \frac{1}{m_i}\,\frac{\partial U}{\partial \zeta_i} \qquad\qquad (i = 1, 2, 3)\,.
\end{aligned}$$

The function U in the three body problem is

$$\frac{m_1m_2}{r_{12}} + \frac{m_1m_3}{r_{13}} + \frac{m_2m_3}{r_{23}}\,,$$

and in the helium atom it is

$$\frac{e_1e_2}{r_{12}} - \frac{e_1e_3}{r_{13}} + \frac{e_2e_3}{r_{23}}\,,$$

e_2 being the charge on the nucleus and $-e_1$, $-e_3$ the charges on the electrons.

These equations can be thrown into the form

$$\frac{dx_i}{dt} = X_i(x)\,, \qquad\qquad (i = 1, \cdots, 18)$$

by the simple transformation

$$\xi_i = x_i\,, \quad \xi_i' = x_{i+3}\,, \quad \eta_i = x_{i+6}\,, \quad \eta_i' = x_{i+9}\,, \quad \zeta_i = x_{i+12}\,, \quad \zeta_i' = x_{i+15}\,.$$

Received March 16, 1935.

[1] Am. Journal of Math., vol. 45 (1923), pp. 93–121; and vol. 50 (1928), pp. 613–626.

[2] Am. Math. Monthly, vol. 38 (1931), pp. 511–521 and vol. 40 (1933), pp. 532–537.

2. Integrals of the equations of variation. We consider a system of differential equations of the form

$$\frac{dx_i}{dt} = X_i\,(x, t) \qquad (i = 1, \cdots, n). \tag{2}$$

In equations (2), and in what follows, when x occurs as an argument it stands for the n-partite variable $x_1, \cdots, x_n$. We consider a particular solution

$$x_i = x_i(t) \tag{3}$$

which is defined for values of t in a region $\mathfrak{T}$ of the complex plane. The functions $X_i(x, t)$ and the functions $F(x, t)$ which will define integrals of equations (2) are assumed to be continuous and to have continuous second partial derivatives in a neighborhood of the elements (x, t) belonging to the solution (3).

The equations of variation for the solution (3) are

$$\frac{d\xi_i}{dt} = \sum_k \frac{\partial}{\partial x_k} X_i(x(t), t)\,\xi_k \qquad (i, k = 1, \cdots, n). \tag{4}$$

If $F(x, t)$ is an integral of the differential equations (2), the function

$$\sum_k \partial/\partial x_k\, F(x(t), t)\xi_k \tag{5}$$

is an integral of the equations of variation (4) for the solution (3) of equations (2).[3]

3. Equilibrium points and their equations of variation. We will consider a system of differential equations of the form

$$\frac{dx_i}{dt} = X_i(x, t) \qquad (i = 1, \cdots, n) \tag{6}$$

in which the functions $X_i(x, t)$ are periodic in t with period τ and vanish identically in t when $x_1 = \cdots = x_n = 0$. The origin is called an equilibrium point for the system. The generalized equilibrium problem[4] defined by a system of this type is associated with motions near a periodic motion $y_i = y_i(t)$ of period τ for the system

$$\frac{dy_i}{dt} = Y_i(y) \tag{7}$$

by means of the transformation $y_i = y_i(t) + x_i$. In case the functions X_i in (6) are independent of t we have the simplest type of equilibrium point and

[3] E. T. Whittaker, *Analytical Dynamics*, 3rd ed., Cambridge, 1927, p. 270.

[4] G. D. Birkhoff, *Dynamical Systems*, Am. Math. Soc., Colloquium Publications, vol. 9 (1927), pp. 59, 60.

equations (6) are identical with the equations (7). The equations of variation for the solution $x_1 = \cdots = x_n = 0$ of equations (6) are

$$\frac{d\xi_i}{dt} = \sum_k \frac{\partial}{\partial x_k} X_i(0, t)\, \xi_k . \tag{8}$$

The coefficients of the variables $\xi_1, \cdots, \xi_n$ either are periodic with period τ or, in the simple case, are constants.

The equations (8) possess a fundamental set of solutions of the form

$$\xi_{ik} = P_{ik}(t)e^{\lambda_k t} \qquad (i, k = 1, \cdots, n),$$

where the numbers $\lambda_1, \cdots, \lambda_n$ are the characteristic exponents for the equations (8). If the coefficients $\partial/\partial x_k X_i(0, t)$ are constants, the functions $P_{ik}(t)$ are polynomials in t of degree not greater than n. If the coefficients in equations (8) are periodic with period τ, the constant coefficients in the polynomials $P_{ik}(t)$ are replaced by periodic functions of t.[5] The general solution of equations (8) can then be written in the form

$$\xi_i = \sum_k c_k\, P_{ik}(t)\, e^{\lambda_k t} , \tag{9}$$

where $c_1, \cdots, c_n$ are arbitrary constants.

Since there is no value of t for which the square matrix $(P_{ik}(t))$ is singular,[6] it has a reciprocal matrix $(Q_{ik}(t))$, also non-singular. Consider the functions

$$\varphi_i(\xi, t) = e^{-\lambda_i t} \sum_j Q_{ij}(t)\, \xi_j \qquad (i, j = 1, \cdots, n) . \tag{10}$$

If we replace the variables ξ_j by the general solutions (9) of the equations of variation and interchange the order of summation we will have

$$\varphi_i(\xi, t) = e^{-\lambda_i t} \sum_k c_k\, e^{\lambda_k t} \Big(\sum_j Q_{ij}\, P_{jk}\Big) = c_i ,$$

since the equations

$$\sum_j Q_{ij}(t)\, P_{jk}(t) = \delta_{ik} ,$$

where $\delta_{ik} = 0$ if $i \neq k$ and $\delta_{ii} = 1$ for every i, hold identically in t. The system of integrals (10) is a fundamental system of integrals since the matrix of coefficients of the variables $\xi_1, \cdots, \xi_n$ in (10) is non-singular for every value of t. Thus we can solve the equations (10) for the variables $\xi_1, \cdots, \xi_n$ and obtain solutions of the form

$$\xi_i = \sum_j \zeta_{ij}(t)\, \varphi_j .$$

If we substitute these variables into any integral $\psi(\xi, t)$ of equations (8) which is linear in the variables $\xi_1, \cdots, \xi_n$, it will appear that ψ is expressed as a linear

[5] F. R. Moulton, *Differential Equations*, Macmillan, 1930, p. 286.

[6] F. R. Moulton, loc. cit., p. 235.

function of the integrals $\varphi_1, \cdots, \varphi_n$ with coefficients which are functions of t. But these coefficients must be constants, for ψ can be expressed as a function of the variables $\varphi_1, \cdots, \varphi_n$ alone.[7] These facts may be summarized in the following theorem.

THEOREM 1. *The system of linear equations* (8) *possesses a fundamental system of integrals of the form* (10). *Every other integral* $\psi(\xi, t)$ *of the equations of variation* (8) *which is linear in the variables* $\xi_1, \cdots, \xi_n$ *may be expressed as a linear form in the variables* $\varphi_1, \cdots, \varphi_n$ *with constant coefficients. The determinant*

$$| e^{-\lambda_k t} Q_{ij}(t) |$$

of the coefficients in (10) *vanishes for no value of* t.

4. **Relations between integrals and the characteristic exponents.** We will prove a group of theorems concerning the manner in which the form of the integrals of a dynamical system (6) may determine the nature of the characteristic exponents for an equilibrium point.

We consider a system of differential equations (6) which have a simple or generalized equilibrium point at the origin and which have p integrals

$$F_s(x, t) \qquad (s = 1, \cdots, p \leqq n), \tag{11}$$

which are such that the $p \times n$ matrix

$$\left(\frac{\partial}{\partial x_k} F_s(0, t)\right) \tag{12}$$

has rank p for every value of t. The linearly independent integrals of the equations of variation (8) which the integrals (11) determine are

$$\sum_k \partial/\partial x_k F_s(0, t)\, \xi_k . \tag{13}$$

THEOREM 2. *If the* p *integrals* (11) *of equations* (6) *are periodic in* t *with period* T, *then* p *characteristic exponents are of the form* $2\pi\nu\sqrt{-1}/T$, *where* ν *is an integer.*

To prove this theorem, we note that the hypotheses insure that there exist p linearly independent integrals of the form (13) of the equations of variation which are periodic with period T. This fact and Theorem 1 demand that p integrals of the fundamental set (10) be periodic with period T. For the i-th integral of the set (10) to be periodic with period T, it is necessary that λ_i be of the form $2\pi\nu\sqrt{-1}/T$, where ν is an integer.

THEOREM 3. *If the* p *integrals* (11) *of equations* (6) *are such that the integrals* (13) *of the equations of variation are rational functions of* t, *then* p *characteristic exponents are zero.*

For Theorem 1 together with the form of the integrals (10) implies that p

[7] F. R. Moulton, *Differential Equations*, pp. 73–75.

integrals of the set (10) must be rational functions of t. Furthermore, if the i-th integral of the set (10) has not an essential singularity at $t = \infty$, we see that in the case of simple equilibrium the coefficient $e^{-\lambda_i t}$ must be a constant, while in the case of generalized equilibrium $e^{-\lambda_i t}$ must be a constant or must be periodic with period τ. In any case we must have $\lambda_i = 0$, since we can and do replace an exponent of the form $2\pi\nu\sqrt{-1}/T$ by a zero exponent in the solutions of the equations of variation for the generalized equilibrium problem.

This theorem includes as a special case the well-known theorem for the case of simple equilibrium that if p integrals (11) are independent of t, then p characteristic exponents are zero.[8] However, in case the right members of the equations of variation (8) are periodic in t, the theorem of Poincaré[9] is not included, since the reduction to generalized equilibrium excludes his argument.

THEOREM 4. *If the equations* (6) *have an integral which is such that the corresponding integral of the equations of variation* (13) *is periodic in* t *with a minimum period* T, *incommensurable with* τ *in the case of generalized equilibrium, then two of the characteristic exponents are* $\pm 2\pi\sqrt{-1}/T$.

Proof. The integral (13) of the equations of variation is periodic with minimum period T and this integral is, by Theorem 1, a linear combination of the integrals (10). If we take into account the form of the integrals (10), we see that one of the integrals (10) must have minimum period T. This implies that one pair of characteristic exponents must be $\pm 2\pi\sqrt{-1}/T$.

THEOREM 5. *If* p *integrals of the equations of variation are continuous and bounded as functions of* t *for all real values of* t, *then* p *characteristic exponents are pure imaginaries.*

The hypotheses of the theorem together with Theorem 1 imply that p of the integrals of the fundamental set (10) are bounded as functions of t for all real values of t. In order that the i-th integral of the set (10) have this property, it is necessary that the real part of λ_i be zero.

One may apply Theorem 3 and Theorems 4 and 5, assuming that the integrals have real periods, to obtain sufficient conditions for stability in the sense that the characteristic exponents are all pure imaginaries and distinct except for possible zero exponents.

5. Applications to the problems of three bodies and the neutral helium atom. Equations (1) have ten integrals. It is desirable to use the six center of gravity integrals to eliminate $\bar{\xi}_2$, $\bar{\eta}_2$ and $\bar{\zeta}_2$. We shall consider that this has been done. There remain the three area integrals and the energy integral. If the equilibrium points are given by

$$\bar{\xi}_i = a_i, \qquad \bar{\eta}_i = b_i, \qquad \bar{\zeta}_i = 0 \qquad (i = 1, 3),$$

[8] H. Poincaré, *Les méthodes nouvelles de la mécanique céleste*, Paris, 1892, vol. 1, p. 192.
[9] Ibid., p. 188.

b_i being zero in the straight line positions, and if we introduce new variables by the equations

$$\xi_i = a_i + x_i, \qquad \eta_i = b_i + y_i, \qquad \zeta_i = z_i,$$

then the integrals take the forms,

$$F_1 \equiv \sum_1^3 m_i \left[\omega\{(a_i + x_i)^2 + (b_i + y_i)^2\} + (a_i + x_i)\frac{dy_i}{dt} - (b_i + y_i)\frac{dx_i}{dt} \right],$$

$$F_2 \equiv \sum_1^3 m_i \left\{ (a_i + x_i)\frac{dz_i}{dt} - z_i\frac{dx_i}{dt} + \omega(b_i + y_i)z_i \right\} \sin \omega t$$

$$+ \sum_1^3 m_i \left\{ (b_i + y_i)\frac{dz_i}{dt} - z_i\frac{dy_i}{dt} - \omega(a_i + x_i)z_i \right\} \cos \omega t,$$

$$F_3 \equiv \sum_1^3 m_i \left\{ z_i\frac{dy_i}{dt} - (b_i + y_i)\frac{dz_i}{dt} + \omega(a_i + x_i)z_i \right\} \sin \omega t$$

$$+ \sum_1^3 m_i \left\{ (a_i + x_i)\frac{dz_i}{dt} - z_i\frac{dx_i}{dt} + \omega z_i(b_i + y_i) \right\} \cos \omega t,$$

$$F_4 = \sum_1^3 m_i \left[\left(\frac{dx_i}{dt}\right)^2 + \left(\frac{dy_i}{dt}\right)^2 + \left(\frac{dz_i}{dt}\right)^2 - \omega^2\{(a_i + x_i)^2 + (b_i + y_i)^2\} \right] - U.$$

The formulas (13) enable us to write at once the integrals of the equations of variation for the particular solutions which give the equilibrium points. The elimination of x_2, y_2 and z_2 does not alter the character of the integrals. Two of these integrals are independent of the time. Hence by Theorem 3 two of the characteristic exponents must be zero, since all the conditions of the theorem are fulfilled. The other two integrals are periodic with minimum period $2\pi/\omega$. All other conditions of Theorem 4 are satisfied, hence two of the characteristic exponents must be $+\omega\sqrt{-1}$ and $-\omega\sqrt{-1}$.

TULANE UNIVERSITY.

SOME IRREDUCIBLE MONOMIAL REPRESENTATIONS OF HYPERORTHOGONAL GROUPS

By J. S. Frame

1. A large number of simple groups of finite order can most easily be defined by means of matrices with coefficients from a finite field, whose characteristic p is a factor of the order of the group. A certain infinite family of these simple groups may be represented by unitary matrices of degree m with coefficients from a finite field $GF(q^2)$ of q^2 elements. Here q is the power p^s of a prime p, and to each number x a conjugate is defined by the relation $\bar{x} = x^q$. Since each x from the $GF(q^2)$ satisfies the equation $x^{q^2} = x$, it follows that $\bar{\bar{x}} = \bar{x}^q = x^{q^2} = x$. By an m-dimensional GF-vector a, we shall mean an ordered set $(a_1, a_2, \cdots, a_m)$ of m numbers from the field $GF(q^2)$, and we shall use the notation $\sum_{i=1}^{m} \bar{a}_i b_i = (a \mid b) = (\overline{b \mid a})$, calling the vectors a and b orthogonal if $(a \mid b) = 0$.

In a recent paper[1] the author has studied some of these simple groups, resolved them into sets of conjugate operations, and found for each group a representation as a permutation group of degree $q^3 + 1$, which was easily reduced into its two irreducible components. In this paper we shall find a set of monomial representations of these groups, also of degree $q^3 + 1$, with complex coefficients, some of which are irreducible, and the rest of which split into two irreducible components. Together these determine more than half of the distinct irreducible representations. With the aid of the familiar relations between characters, we are then able to determine the degrees and most of the characters of all the irreducible representations of these groups.

Let G_m^* be the group of unitary matrices of degree m in this Galois field $GF(q^2)$; that is, the group of those matrices T which leave invariant the form $(x \mid x)$. The matrices have the elements (t_{ij}), where $\sum_{k=1}^{m} \bar{t}_{ik}\, t_{jk} = \sum_{k=1}^{m} \bar{t}_{ki}\, t_{kj} = \delta_{ij}$. In short, T is the transposed matrix of $\overline{T}$. Since its determinant satisfies the equation $T\overline{T} = 1$, it can have one of only $q + 1$ possible values. If we write $a' = aT$ when $a'_j = \sum_{i=1}^{m} a_i t_{ij}$, $(j = 1, 2, \cdots, m)$, we find $(aT \mid bT) = (a \mid b)$ for every matrix T of G_m^*, and for all GF-vectors a and b. Dickson[2] has shown that the order of G_m^* is

Received March 22, 1935.

[1] J. S. Frame, *Unitäre Matrizen in Galoisfeldern*, Commentarii Mathematici Helvetici, vol. 7 (1935), p. 94. This paper will be cited here as U.

[2] L. E. Dickson, *Linear Groups*, 1907.

$$g_m^* = \prod_{k=1}^{m} q^{k-1}Q_k = q^{\binom{m}{2}} \prod_{k=1}^{m} Q_k, \text{ if we write } Q_k = q^k - (-1)^k.$$

The matrices of determinant unity form an invariant subgroup G_m of index Q_1 under G_m^*, and of order $g_m = g_m^*/Q_1$. The central K_m of this subgroup consists of the multiplications $x_i' = \alpha x_i$, such that $\alpha\bar{\alpha} = 1$, and $\alpha^m = 1$. There are just d such, if d is the h.c.f. of m and $q + 1$. Dickson has shown that the quotient group $H_m = G_m/K_m$ of order $h_m = g_m/d$, a group which he calls hyperorthogonal and denotes by $HO(m, p^{2s})$, is simple except in the three cases $HO(2, 4)$, $HO(2, 9)$, $HO(3, 4)$. He has proved further that $HO(2, q^2)$ is isomorphic with the linear fractional group $LF(2, q)$ of order $q(q^2 - 1)/d$. These linear fractional groups are quite well known, so we pass them by to study especially the groups $HO(3, q^2)$, which for $q = 3, 4, 5, 7, 8$ have the orders 6048, 62400, 126000, 5663616, 5515776, respectively.

2. We shall first exhibit certain normal forms which display quite simply the matrices of some of the complete sets of conjugate operations of G_m. In U it was shown that the matrices

(I) $M(a; \epsilon) \equiv (\delta_{ij} + \epsilon\bar{a}_i a_j)$, where $\epsilon + \bar{\epsilon} = (a \mid a) = 0$, but $\epsilon a \neq 0$, form a single complete set of conjugate operations of order p in G_m, satisfying the relations

(I′) $M(a; \epsilon_1)\cdot M(a; \epsilon_2) = M(a; \epsilon_1 + \epsilon_2)$,

(I″) $T^{-1}M(a; \epsilon)T = M(aT; \epsilon)$, for T in G_m.

It was shown also that the matrices

(II) $M(a; \alpha; \epsilon) \equiv (\delta_{ij} + \bar{\alpha}_i a_j - \bar{a}_i \alpha_j + \epsilon\bar{a}_i a_j)$, where $(a \mid a) = (\alpha \mid a) = (\alpha \mid \alpha) + \epsilon + \bar{\epsilon} = 0$, are of order p, if $p > 2$, or p^2, if $p = 2$, and form one or d complete sets of conjugates, according as $m > 3$ or $m = 3$.

THEOREM 1. *The matrices*

(III) $M'(c; \theta) \equiv (\theta\delta_{ij} + \theta(\theta^{-m} - 1)\bar{c}_i c_j/(c \mid c))$, *where* $(c \mid c) \neq 0, \theta\bar{\theta} = 1, \theta \neq 1$, *form for each admissible value of* θ *a single set of* $q^{m-1}Q_m/Q_1$ *conjugate operations of* G_m, *whose orders divide* Q_1. *They satisfy the relations*

(III′) $M'(c; \theta_1)\cdot M'(c; \theta_2) = M'(c; \theta_1\theta_2)$,

$M'(kc; \theta) = M'(c; \theta)$, *for* $k \neq 0$,

and

(III″) $T^{-1}M'(c; \theta)T = M'(cT; \theta)$, *for* T *in* G_m.

The proofs of III′ and III″ follow immediately from the multiplication of matrices. The matrix becomes a diagonal matrix if all but one of the components of c are made to vanish; its determinant is then readily seen to be unity. Since the character, or trace, of $M'(c; \theta)$ is $\theta(m - 1) + \theta^{1-m}$, two such matrices can be identical only if they have the same value of θ, $(\theta \neq 1)$, and if the vectors c are proportional. By Theorem 5 in U there are $Q_m Q_{m-1}$ vectors a, $a \neq 0$, for which $(a \mid a) = 0$. Hence there are $q^{2m} - 1 - Q_m Q_{m-1} = Q_m(q^m - q^{m-1})$ vectors c for which $(c \mid c) \neq 0$, and $q^{m-1}Q_m/Q_1$ matrices for each θ. The fact that these are all conjugate to each other depends on the fact (see U) that the vectors c are permuted transitively among themselves by the matrices T of G_m.

We find that $M'(c; \theta_1)$ and $M'(c'; \theta_2)$ are permutable if and only if either $c = kc'$ or $(c \mid c') = 0$, and that $M(a; \epsilon)$ and $M'(c; \theta)$ are permutable if and only if $(a \mid c) = 0$. Now a matrix of order $p(q + 1)$ will be obtained by multiplying together two permutable matrices of order p and $q + 1$, respectively, which will exist only if $m > 2$.

THEOREM 2. *The matrices*

(IV) $M'(a; \epsilon; c, \theta) \equiv (\theta\delta_{ij} + \theta\epsilon\bar{a}_i a_j + \theta(\theta^{-m} - 1)\bar{c}_i c_j/(c \mid c))$, *where* $\epsilon + \bar{\epsilon} = (a \mid a) = (c \mid a) = 0$, $\epsilon a \neq 0$, $(c \mid c) \neq 0$, $\theta\bar{\theta} = 1$, $\theta \neq 1$, *form for each admissible* θ *a single set of* $q^{m-1}Q_m Q_{m-1} Q_{m-2}/Q_1^2$ *conjugate operations in* G_m, *whose orders divide* $p(q + 1)$. *They satisfy the relation*

(IV″) $T^{-1}M'(a; \epsilon; c, \theta)T = M'(aT; \epsilon; cT, \theta)$.

The proof hinges on the fact, (see U), that the subgroup of G_m which leaves a vector c fixed permutes transitively among themselves the vectors a orthogonal to c, at least to within a multiplicative factor which does not affect the matrix. This shows that these matrices are all conjugate to each other in G_m. To determine their number, we find that $Q_{m-1}Q_{m-2}/Q_1$ matrices $M(a; \epsilon)$ are permutable with each of the $q^{m-1}Q_m/Q_1$ matrices $M'(c; \theta)$, and that a matrix $M'(a; \epsilon; c, \theta)$ can be resolved in only one way, namely, $M = M^{q+1} \cdot M^{-q}$, into a product of the form $M(a; \epsilon) \cdot M'(c; \theta)$.

3. We turn for a moment from an algebraic to a geometric point of view. In U it was shown that there are $Q_m Q_{m-1}$ GF-vectors $a \neq 0$ for which $(a \mid a) = 0$, and that corresponding to these vectors there are $Q_m Q_{m-1}/Q_2$ rays R_a, each consisting of the Q_2 non-zero multiples of a given vector a. It was also shown that these rays are permuted transitively among themselves by the matrices of G_m, when $m > 2$, and further that the subgroup of G_m leaving one ray R_a invariant permutes transitively among themselves the rays R_c for which $(a \mid c) \neq 0$, and also among themselves the rays $R_b \neq R_a$ for which $(a \mid b) = 0$. This led to a representation of the group H_m as a transitive permutation group P_m on $Q_m Q_{m-1}/Q_2$ symbols. Considered as a group of linear transformations it was shown that this group has just three irreducible components for $m > 3$, and only two for $m = 3$. (This exception is due to the fact that when $m = 3$ there are no rays $R_b \neq R_a$ such that $(a \mid b) = 0$.) Of these two components for $m = 3$, one is the identity representation, and the other is an irreducible representation of degree q^3.

By altering slightly the form of the permutation group P_m, we now find a set of monomial representations of H_m. From each ray R_a we pick a particular vector a. Now since the rays are permuted transitively by the permutation group P_m, *these vectors undergo a monomial representation in which the factors are marks from the* $GF(q^2)$, satisfying the equation $x^{q^2-1} = 1$. Since in products of these monomial matrices only the multiplicative property of the factors comes into play, *we may replace these factors by ordinary* $(q^2 - 1)$*-th roots of unity in the field of complex numbers*. Thus we obtain $q^2 - 1$ monomial representations of G_m, some of which may be equivalent. Now H_m is the factor group of G_m, of

index d, obtained by letting the transformations $x'_i = \alpha x_i$, where $\alpha^d = 1$, correspond to identity. So if in all these monomial representations of G_m we replace each multiplier by its d-th power, then each element of the central of G_m, and these alone, will be represented by the identity, and we shall obtain a set of isomorphic representations of H_m.

The questions of reducibility and equivalence of these representations may be answered by examining their characters. For any given matrix the trace, or character, is the sum of the multipliers affecting the invariant rays. (This sum is now to be interpreted in the field of complex numbers.) Conjugate matrices have the same trace, and two isomorphic representations are equivalent if and only if corresponding matrices have the same traces. The sum, taken over all the matrices of a group, of the squared absolute value of the trace is a multiple of the order of the group, this multiple being unity if and only if the group is irreducible.

One or two restrictions on the multipliers of these matrices are of importance in determining the characters. The properties of transitivity mentioned above, (when $m > 2$), show that corresponding to a matrix T leaving one ray, or two non-orthogonal rays, or two mutually orthogonal rays invariant, there is some conjugate matrix which does the same to arbitrary rays with the same properties of orthogonality. If the matrix T takes the vectors a_0 into k_1a_0, and b_0 into k_2b_0, then $(a_0 \mid b_0) = (k_1a_0 \mid k_2b_0) = \bar{k}_1k_2(a_0 \mid b_0)$. If $(a_0 \mid b_0) \neq 0$, we must have $\bar{k}_1k_2 = 1$, which means $k_2 = k_1^{-q}$. If a third ray R_{c_0} is invariant under T, (i.e., if $c_0T = k_3c_0$) and if $(a_0 \mid c_0) \neq 0$, $(b_0 \mid c_0) \neq 0$, it follows that $k_1 = k_2 = k_3 = \theta$, where $\theta\bar{\theta} = 1$. In this case all the linear combinations of a_0, b_0, c_0 are multiplied by the same factor θ, so the corresponding rays are all invariant. Now for $m > 2$ we can always find matrices which leave two non-orthogonal rays, and no others, relatively invariant. Let the complex numbers ϕ and ϕ^{-q}, where $\phi^{Q_2/d} = 1$, correspond to these G.F. factors k_1, and $k_2 = k_1^{-q}$, which affect the rays. Then for $(q-2)Q_1/2d$ values of $\phi \neq \phi^{-q}$, and for Q_1/d values of $\phi = \phi^{-q}$, the characters $\phi + \phi^{-q}$ are all different. Hence *we have obtained in this way at least $qQ_1/2d$ distinct monomial representations of $H_m(m > 2)$.*

4. We are now ready to study in detail the characters of the monomial representations of H_m when $m = 3$. In this case no two distinct rays R_a with $(a \mid a) = 0$ are orthogonal, so the above restrictions on the multipliers are valid. (1) Only the d matrices of the central of G_m—those which correspond to the identity in H_m—can leave three linearly independent rays invariant, and they will leave each of $q^3 + 1$ rays absolutely invariant. Hence the identity has the character $q^3 + 1$ in each of the monomial representations. (2) In addition to the identity there are $(Q_1/d) - 1$ sets of $\binom{Q_3}{2} \Big/ \binom{Q_1}{2} = q^2Q_3/Q_1$ conjugate elements of the form III discussed in §2. These multiply each of the Q_1 linear combinations of two linearly independent vectors by the same factor θ^n. In the monomial representations these have the character $Q_1(\phi^{q-1})^n$, where ϕ is a complex number

satisfying the equation $\phi^{Q_2/d} = 1$, and $n = 1, 2, \cdots, (Q_1/d) - 1$. (3) In like manner we find $(q-2)Q_1/2d$ sets of Q_3q^3 conjugate elements in H_3 which affect two vectors a_0 and b_0 with different factors, and have the character $\phi^n + \phi^{-nq}$, as given above. The $(q-1)$-th power of such a matrix lies in one of the sets (2) just discussed, since it multiplies both vectors by the same factor.

Matrices leaving just one ray R_a invariant may either leave it absolutely invariant or multiply it by a factor. They are all included in the subgroup of $h_3/(Q_3Q_2/Q_1) = q^3Q_1/d$ matrices permutable with the matrix $M(a; \epsilon)$. Hence their orders are factors of q^3Q_1/d. (4) The matrices $M(a; \epsilon)$, of the form (I) discussed in §2, form a single set of Q_3Q_2/Q_1 conjugates, each of order p, and correspond to matrices with character 1 in the monomial representation. (5) The matrices (II) of order 4 (when $p = 2$) or p (when $p > 2$) form d sets of Q_3Q_2q/d elements, also corresponding to monomial matrices with character 1. (6) The matrices (IV), whose orders divide pQ_1/d, cannot leave a ray absolutely invariant, since their p-th powers do not. But since the Q_1/d-th power of each of these does leave a ray invariant, the corresponding multiplier must be of the form $(\phi^{q-1})^n$, $n = 1, 2, \cdots, (Q_1/d) - 1$. Here we have $(Q_1/d) - 1$ sets, each of $Q_3Q_2q^2/Q_1$ conjugate monomial matrices.

There remain two or three types of sets of conjugate matrices which permute the $q^3 + 1$ rays, leaving none invariant, and have therefore the character 0 in the monomial representation. Their orders must divide $q^3 + 1$. Of this type are the diagonal matrices with three distinct multipliers θ^i, θ^j, θ^{-i-j}, where $\theta\bar{\theta} = 1$. Such triples may be chosen in $(Q_1^2 - 3Q_1 + 2d)/6$ distinct ways. When $d = 3$, triples obtained from one another by multiplication by ω or ω^2, where $\omega^3 = 1$, correspond to the same element of H_3, so that in general d sets of conjugates in G_3 collapse into one in H_3. Only a matrix of order d, such as the diagonal matrix with multipliers $1, \omega, \omega^2$, is taken into a conjugate matrix in G_3 by such a multiplication. It is permutable, up to a factor ω^i, with that subgroup of H_3 of order Q_1^2 generated by the $Q_1^2/3$ multiplications and a permutation of order 3. This gives us (7): a set of h_3/Q_1^2 conjugates of order d, when $d = 3$ (otherwise like type (8)). The rest of these matrices of diagonal type are permutable only with the group of Q_1^2/d diagonal matrices, and (8) they belong to $(q^2 - q - 8 + 2d)/6d$ sets of h_3d/Q_1^2 conjugate matrices in H_3, whose orders divide Q_1/d.

The final type (9) to be considered includes matrices of order $(q^2 - q + 1)/d$. Since for the character of any matrix T in the $GF(q^2)$ representation we have $\chi(T^{q^2}) = \chi(T)$, each matrix of order $(q^2 - q + 1)/d$ has the same character as its q^2-th and q^4-th powers—that is, as its $(q-1)$-th and $(-q)$-th powers —and can be shown to be conjugate to them. The identity excluded, the $(q^2-q+1-d)/d$ powers of one properly chosen matrix lie in $(q^2-q+1-d)/3d$ sets of h_3Q_1d/Q_3 conjugates, since each such matrix is permutable only with a cyclic group of order $(q^2 - q + 1)/d$. All the h_3 transformations of H_3 are now accounted for. We have a total of $q(q + 1)/d + d + 1$ sets of conjugates, so this must be the number of irreducible representations of the group H_3.

In summary, we tabulate in four columns the number of sets of each of the

nine types, the number of matrices in each set, the character of a matrix of the set in the monomial representation corresponding to a particular $(q^2 - 1)/d$-th root of unity ϕ, and a common multiple of the orders of matrices of that type.

Sets	Matrices in set	Character	Multiple of order
1	1	$q^3 + 1$	1
$(q+1)/d - 1$	$(q^2 - q + 1)q^2$	$(q+1)(\phi^{q-1})^n$	$(q+1)/d$
$(q^2 - q - 2)/2d$	$(q^3 + 1)q^3$	$\phi^n + \phi^{-nq}$	$(q^2 - 1)/d$
1	$(q^3 + 1)(q - 1)$	1	p
d	$(q^3 + 1)(q^2 - 1)q/d$	1	p, or $4 = p^2$.
$(q+1)/d - 1$	$(q^3 + 1)(q - 1)q^2$	$(\phi^{q-1})^n$	$p(q+1)/d$
1	$(q^2 - q + 1)(q - 1)q^3/d$	0	$(q+1)$
$(q^2 - q - 8 + 2d)/6d$	$(q^2 - q + 1)(q - 1)q^3$	0	$(q+1)$
$(q^2 - q + 1 - d)/3d$	$(q^2 - 1)(q + 1)q^3$	0	$(q^2 - q + 1)/d$

5. The standard relations between characters show that the $(q^2 - q - 2)/2d$ monomial representations obtained from values of $\phi \neq \phi^{-q}$ are all irreducible, while the others with $\phi = \phi^{-q}$ each have exactly two irreducible components. The permutation group P_3 discussed in U is obtained by taking $\phi = 1$, and is easily split into its two irreducible components. The problem of splitting the others is not so easy. First we must write the degree $q^3 + 1$ as the sum of two factors of the order $h_3 = (q^3 + 1)(q^2 - 1)q^3/d$. One of these must be either 1, $q + 1$, or $q^2 - q + 1$. The first gives the identity component of P_3 (no other representations of degree 1 are possible in a simple group), but the second is found to involve incompatible conditions on the characters of the elements of order $(q^2 - q + 1)/d$. There remains only the possibility $(q^3 + 1) = q(q^2 - q + 1) + (q^2 - q + 1)$. By actually working out the tables of characters for $q = 3, 4, 5$, with the aid of many known relations between characters of irreducible representations, it was found possible to obtain for most of these characters literal expressions in terms of q and d, such as to satisfy the required relations in literal form. With these at hand, we then constructed the tables for $q = 7$ and 8, and obtained further general information about those characters whose literal form was not yet determined. Omitting further details, we note in conclusion that the $q(q + 1)/d + d + 1$ distinct irreducible representations fall into nine types, as did the sets of conjugates, and that their degrees are as follows.

1	of	degree	1
1	"	"	q^3
1	"	"	$q^2 - q$
$(q^2 - q + 1 - d)/3d$	"	"	$(q^2 - 1)(q + 1)$
d	"	"	$(q^2 - q + 1)(q - 1)/d$
$(q^2 - q - 8 + 2d)/6d$	"	"	$(q^2 - q + 1)(q - 1)$
$(q^2 - q - 2)/2d$	"	"	$q^3 + 1$
$((q + 1)/d) - 1$	"	"	$q(q^2 - q + 1)$
$((q + 1)/d) - 1$	"	"	$q^2 - q + 1$

To illustrate the theory, we give below the complete table of characters for the group $HO(3, 25)$, of order 126000. Here we have $q = 5$, and $d = 3$. Each row corresponds to a complete set of h_λ conjugates, and each column to an irreducible representation. The numbers h_λ are on the left, and h/h_λ on the right.

h_λ															h/h_λ
1	1	125	20	144	144	28	28	28	84	126	126	126	105	21	126000
525	1	5	-4	0	0	-4	-4	-4	4	-6	6	-6	1	5	240
15750	1	1	0	0	0	0	0	0	0	$i\sqrt{2}$	0	$-i\sqrt{2}$	-1	-1	8
15750	1	1	0	0	0	0	0	0	0	0	-2	0	1	1	8
15750	1	1	0	0	0	0	0	0	0	$-i\sqrt{2}$	0	$i\sqrt{2}$	-1	-1	8
504	1	0	-5	-6	-6	3	3	3	9	1	1	1	5	-4	250
5040	1	0	0	-1	-1	3	-2	-2	-1	1	1	1	0	1	25
5040	1	0	0	-1	-1	-2	3	-2	-1	1	1	1	0	1	25
5040	1	0	0	-1	-1	-2	-2	3	-1	1	1	1	0	1	25
12600	1	0	1	0	0	1	1	1	-1	-1	1	-1	1	0	10
3500	1	-1	2	0	0	1	1	1	3	0	0	0	-3	3	36
10500	1	-1	2	0	0	-1	-1	-1	1	0	0	0	1	-1	12
18000	1	-1	-1	$\frac{1}{2}(1 + i\sqrt{7})$	$\frac{1}{2}(1 - i\sqrt{7})$	0	0	0	0	0	0	0	0	0	7
18000	1	-1	-1	$\frac{1}{2}(1 - i\sqrt{7})$	$\frac{1}{2}(1 + i\sqrt{7})$	0	0	0	0	0	0	0	0	0	7

BROWN UNIVERSITY.

NECESSARY AND SUFFICIENT CONDITIONS IN THE MOMENT PROBLEM FOR A FINITE INTERVAL

By R. P. Boas, Jr.

1. **Introduction.** The moment problem of F. Hausdorff is the determination of necessary and sufficient conditions that a sequence of numbers $\{\mu_n\}$ have the form

$$\int_0^1 t^n d\alpha(t) = \mu_n \qquad (n = 0, 1, 2, \cdots), \tag{1}$$

where $\alpha(t)$ is required to belong to some particular class of functions. If $\alpha(t)$ is an integral, (1) becomes

$$\int_0^1 t^n \varphi(t) dt = \mu_n \qquad (n = 0, 1, 2, \cdots). \tag{2}$$

Since for any $\varphi(t)$ which is integrable (in the sense of Lebesgue) the numbers μ_n, if of the form (2), must have the property that $\mu_\infty = \lim_{n\to\infty} \mu_n = 0$, we shall consider also conditions that a sequence $\{\mu_n\}$ have the form

$$\int_0^1 t^n \varphi(t) dt = \mu_n - \mu_\infty \qquad (n = 0, 1, 2, \cdots). \tag{3}$$

Stating the problem in the form (3) merely serves to simplify the formulation of some of our results.

Hausdorff[1] has obtained necessary and sufficient conditions for the existence of solutions of (1) and (2) under a variety of conditions on $\alpha(t)$ and $\varphi(t)$. Hildebrandt[2] has obtained one of Hausdorff's conditions for the moment problem (1) by utilizing the theory of linear operations,[2a] the polynomials of S. Bernstein, and a classical theorem of F. Riesz on the general form of a linear functional in the space of continuous functions. Professor Widder suggested to me that it should be possible to obtain conditions analogous to Hausdorff's by the method of Hildebrandt, but using, instead of the Bernstein polynomials, two inversion operators which he has developed[3] for moment sequences known to have the

Received April 2, 1935. The author is indebted to Professor D. V. Widder for suggesting the subject of this paper, and for valuable advice and encouragement during its preparation.

[1] F. Hausdorff, *Momentprobleme für ein endliches Intervall*, Mathematische Zeitschrift, vol. 16 (1923), pp. 220–248.

[2] T. H. Hildebrandt, *On the moment problem for a finite interval*, Bulletin of the American Mathematical Society, vol. 38 (1932), pp. 269–270.

[2a] By a linear operation we shall understand an additive and continuous operation.

[3] D. V. Widder, *The inversion of the Laplace integral and the related moment problem*, Transactions of the American Mathematical Society, vol. 36 (1934), pp. 107–200.

forms (1) and (2). This has been found to be the case; we shall derive below criteria for several forms of the problem treated by Hausdorff, and for some not considered by him. In addition, the method of proof has been considerably simplified by the use, not of the general theorems of Riesz and Steinhaus on the form of linear functionals in various spaces, but of the specialized theorems (of course derivable from the more general results) which give necessary and sufficient conditions for the solvability of the moment problem in certain cases, but which are inconvenient in practice (they are stated below as Lemmas 1, 2, and 3). In this way the continuity argument which is always involved no longer appears explicitly; it is made once for all in the establishment of the general theorems.

The central idea of the method is based on the close relation between the moment problem and the theory of linear operations. For the purpose of concrete illustration we restrict ourselves for the present to the moment problem (1) under the condition that $\alpha(t)$ is a function of bounded variation. If we define an operation $M[P_n(t)]$ on polynomials $P_n(t)$ by the relation

$$(4) \qquad M[P_n(t)] \equiv M\left[\sum_{i=0}^{n} a_i t^i\right] = \sum_{i=0}^{n} a_i \mu_i ,$$

the problem of determining when $\alpha(t)$ of bounded variation exists satisfying (1) is the same as the problem of determining a sufficient condition for the existence of such a function satisfying

$$(5) \qquad \int_0^1 P_n(t) d\alpha(t) = M[P_n(t)]$$

for all polynomials. Now, the necessary and sufficient condition of Riesz for the moment problem (1) to have a solution becomes in this case that for some $K > 0$, and for every polynomial $P_n(t)$,

$$(6) \qquad | M[P_n(t)] | \leqq K \| P_n \| = K \max_{0 \leqq t \leqq 1} | P_n(t) | .$$

Hence, to show that a condition on the sequence $\{\mu_n\}$ is sufficient for the existence of a solution $\alpha(t)$ of (1), of bounded variation, we need only show that it implies the relation (6). We may expect to be able to establish such an implication if we can obtain some function of the μ_n and $P_n(t)$ which approximates $M[P_n(t)]$ arbitrarily closely for every polynomial. Such a function will be obtained in terms of the inversion operator mentioned above.

This inversion operator[4] is defined as follows.[5]

$$L_{k,t}\{\mu_n\} = \frac{(n+k+1)!}{n!k!}(-1)^k \Delta^k \mu_n , \qquad n = \left[\frac{kt}{1-t}\right], \qquad 0 \leqq t < 1 ;$$

$$L_{k,1}\{\mu_n\} = L_{k,1-}\{\mu_n\} .$$

[4] D. V. Widder, op. cit., p. 178.

[5] Here and henceforward, $[x]$ denotes the greatest integer $\leqq x$.

We shall also need another operator defined by Widder; we give its definition here for future reference:

$$S_{k,t}\{\mu_n\} = -\mu_\infty - \sum_{i=n+1}^{\infty} \binom{i+k}{k} (-1)^{k+1}\Delta^{k+1}\mu_i, \qquad n = \left[\frac{kt}{1-t}\right];$$

$$S_{k,0} = -\mu_0, \qquad S_{k,1} = -\mu_\infty .$$

In terms of the operator $L_{k,t}\{\mu_n\}$, a necessary and sufficient condition for (1) to have a solution of bounded variation on $0 \leqq t \leqq 1$ is that for some $K > 0$,

$$\text{(A)} \qquad \int_0^1 | L_{k,t}\{\mu_n\} | \, dt < K \qquad (k = 0, 1, 2, \cdots).$$

This condition has been previously obtained by Widder[6] by a method different from that by which we shall obtain it here.

We shall also obtain, in terms of the same operator, and by the same method, but using different general theorems on the moment problem, several criteria for particular cases of the problem (3). Our principal results, in addition to (A), above, are the following.

For (3) to have a solution $\varphi(t)$ which is, on (0, 1): (B), in the class[7] $L^{(r)}$ ($r > 1$); (C), bounded almost everywhere; (D), in the class $L^{(1)}$; (E), of bounded variation; (F), continuous for $0 \leqq t \leqq 1$; it is necessary and sufficient, respectively, that

$$\text{(B)} \qquad \int_0^1 | L_{k,t}\{\mu_n\} |^r \, dt < K^r \qquad (k = 0, 1, 2, \cdots);$$

$$\text{(C)} \qquad | L_{k,t}\{\mu_n\} | < K \quad (k = 0, 1, 2, \cdots; 0 \leqq t \leqq 1);$$

$$\text{(D)} \qquad \lim_{k,k' \to \infty} \int_0^1 | L_{k,t}\{\mu_n\} - L_{k',t}\{\mu_n\} | \, dt = 0;^{8}$$

$$\text{(E)} \qquad \int_0^1 | dL_{k,t}\{\mu_n\} | < K \qquad (k = 0, 1, 2, \cdots);$$

(F) The sequence of functions $L_{k,t}\{\mu_n\}$ converges uniformly, $0 \leqq t \leqq 1$.

The extremely symmetric form of these criteria makes them analogous to the integral criteria of Hausdorff,[9] which involve Legendre polynomials. Our criteria, it should be noticed, can be expressed solely in terms of the differences of the sequence $\{\mu_n\}$. It can be shown directly, by algebraic manipulation, that the condition (A) is equivalent to the condition given by Hausdorff for the corresponding case; this question will be discussed in the last section of

[6] D. V. Widder, op. cit., p. 194.

[7] As usual, $L^{(r)}$ denotes the class of functions integrable, together with their r^{th} powers, in the sense of Lebesgue.

[8] That is, the sequence $L_{k,t}\{\mu_n\}$ converges in the mean (of order 1).

[9] F. Hausdorff, op. cit., p. 246.

the paper. Condition (C) is, except for notation, precisely Hausdorff's condition. The criteria which have been given by Hausdorff for case (D) all involve Legendre polynomials; a criterion involving merely the differences of the sequence $\{\mu_n\}$ does not appear to have been previously given. The case (F) does not appear to have been treated before. As a corollary of (F), we shall obtain also a condition for the solvability of the moment problem (1) by a continuous function $\alpha(t)$.

2. **Preliminary theorems.** We shall need a number of lemmas, most of which are known results which we merely state with appropriate references. There are first the theorems of F. Riesz and H. Steinhaus referred to in the introduction.

LEMMA 1.[10] *For* (1) *to have a solution $\alpha(t)$ of bounded variation, it is necessary and sufficient that for some $K > 0$ and for every polynomial $P_n(t)$,*

$$(7) \qquad | M[P_n(t)] | \leqq K \max_{0 \leqq t \leqq 1} | P_n(t) | \equiv K \| P_n(t) \|_A .$$

Here $M[P_n(t)]$ is the operator defined by (4).

LEMMA 2.[11] *For* (2) *to have a solution $\varphi(t)$ in $L^{(r)}$ $(r > 1)$, it is necessary and sufficient that for some $K > 0$ and for every polynomial $P_n(t)$,*

$$(8) \qquad | M[P_n(t)] | \leqq K \left(\int_0^1 | P_n(t) |^s \, dt \right)^{\frac{1}{s}} \equiv K \cdot \| P_n \|_B , \qquad \left(\frac{1}{r} + \frac{1}{s} = 1 \right).$$

LEMMA 3.[12] *For* (2) *to have a solution $\varphi(t)$ which is bounded almost everywhere, it is necessary and sufficient that for some $K > 0$ and for every polynomial $P_n(t)$,*

$$(9) \qquad | M[P_n(t)] | \leqq K \int_0^1 | P_n(t) | \, dt \equiv K \| P_n \|_C .$$

We shall also need three inversion theorems for moment sequences which are known to have the forms (1) and (2).

LEMMA 4.[13] *If $\alpha(t)$ is of bounded variation on* $(0, 1)$, *if* $\alpha(1) = 0$, *and if*

$$\mu_n = \int_0^1 t^n \, d\alpha(t) \qquad (n = 0, 1, 2, \cdots),$$

then

$$\frac{\alpha(t+) + \alpha(t-)}{2} = \lim_{k \to \infty} S_{k,t}\{\mu_n\} \qquad (0 < t < 1).$$

[10] F. Riesz, *Sur certains systèmes singuliers d'équations intégrales*, Annales scientifiques de l'École Normale Supérieure, (3), vol. 28 (1911), pp. 33–62; p. 43.

[11] F. Riesz, *Untersuchungen über Systeme integrierbarer Funktionen*, Mathematische Annalen, vol. 69 (1910), pp. 449–497; p. 469.

[12] H. Steinhaus, *Additive und stetige Funktionaloperationen*, Mathematische Zeitschrift, vol. 5 (1919), pp. 186–221. Steinhaus does not actually state this theorem, but it is a simple consequence of his results, and has been explicitly formulated in the form given here by S. Banach, *Théorie des Opérations Linéaires*, 1932, p. 75.

[13] D. V. Widder, op. cit., p. 179.

LEMMA 5.[14] *If $\varphi(t)$ is integrable on* $(0, 1)$, *and if*

$$\mu_n = \int_0^1 t^n \varphi(t)\, dt \qquad (n = 0, 1, 2, \cdots),$$

then

$$\varphi(t) = \lim_{k\to\infty} L_{k,t}\{\mu_n\}$$

almost everywhere on $(0, 1)$.

LEMMA 6. *Under the hypotheses of Lemma 5, if in addition $\varphi(t)$ is continuous for $0 \leqq t \leqq 1$, then*

$$\varphi(t) = \lim_{k\to\infty} L_{k,t}\{\mu_n\}$$

uniformly for $0 \leqq t \leqq 1$.

More generally, it can be shown by methods parallel to those used in the treatment of the corresponding theorem on Laplace integrals,[15] that if $\varphi(t)$ is merely continuous for $a \leqq t \leqq b$, where $0 \leqq a < b \leqq 1$, then

$$\varphi(t) = \lim_{k\to\infty} L_{k,t}\{\mu_n\}$$

uniformly for $a' \leqq t \leqq b'$, where $a < a' < b' < b$. However, we shall not need this more general result.

We proceed to establish Lemma 6. We begin with $\varphi(t)$, a polynomial. It is clearly sufficient to consider the case $\varphi(t) = t^r$, r an integer. Then, for $0 \leqq t < 1$,

$$\begin{aligned} L_{k,t}\{\mu_n\} &= \frac{(n+k+1)!}{n!\,k!} \int_0^1 y^{n+r}(1-y)^k\, dy, \qquad n = \left[\frac{kt}{1-t}\right] \\ &= \frac{(r+n)(r+n-1)\cdots(n+1)}{(r+n+k+1)\cdots(n+k+2)}, \end{aligned}$$

by use of the Beta function. This fraction is the product of r terms of the form

$$\frac{r+n-j}{r+n+k+1-j} = \frac{r + \left[\dfrac{kt}{1-t}\right] - j}{r - j + 1 + k + \left[\dfrac{kt}{1-t}\right]} \qquad (0 \leqq j \leqq r-1).$$

It will be sufficient to show that the limit of each such term is t, uniformly for $0 \leqq t < 1$, as $k \to \infty$. We have

$$\frac{r - j - 1 + \dfrac{kt}{1-t}}{r - j + 1 + k + \dfrac{kt}{1-t}} \leqq \frac{r - j + \left[\dfrac{kt}{1-t}\right]}{r - j + 1 + k + \left[\dfrac{kt}{1-t}\right]} \leqq \frac{r - j + \dfrac{kt}{1-t}}{r - j + k + \dfrac{kt}{1-t}},$$

and it can be seen at once that the two extreme members of this inequality both approach t uniformly on $0 \leqq t < 1$. Hence the middle term does so also. On

[14] D. V. Widder, op. cit., p. 183.

[15] D. V. Widder, op. cit., p. 126.

the other hand, the representation above for $L_{k,t}\{\mu_n\}$ shows that $L_{k,1-}\{\mu_n\} = 1 = t^r]_{t=1} \to 1$; hence the limit is uniformly approached in the closed interval. This establishes the lemma for polynomials.

Now if $\varphi(t)$ is any continuous function, given any $\epsilon > 0$ we determine a polynomial $\psi(t)$ such that

$$|\varphi(t) - \psi(t)| < \epsilon \qquad (0 \leqq t \leqq 1).$$

Set

$$\nu_n = \int_0^1 t^n \psi(t)\, dt \qquad (n = 0, 1, 2, \cdots).$$

Then

$$|L_{k,t}\{\mu_n\} - L_{k,t}\{\nu_n\}| = \frac{(n+k+1)!}{n!\,k!}\left|\int_0^1 y^n(1-y)^k\{\varphi(y) - \psi(y)\}\, dy\right|$$
$$\leqq \epsilon\,\frac{(n+k+1)!}{n!\,k!}\int_0^1 y^n(1-y)^k\, dy = \epsilon \qquad (0 \leqq t < 1).$$

This holds for $0 \leqq t \leqq 1$, by continuity. We then have

$$|\varphi(t) - L_{k,t}\{\mu_n\}| \leqq |\varphi(t) - \psi(t)| + |\psi(t) - L_{k,t}\{\nu_n\}| + |L_{k,t}\{\nu_n\} - L_{k,t}\{\mu_n\}|$$
$$\leqq 3\epsilon,$$

for k sufficiently large, uniformly for $0 \leqq t \leqq 1$. This establishes the lemma.

We shall need one general result on sequences $\{\mu_n\}$.

Lemma 7.[16] *If*

$$\left|\sum_{p=m}^{\infty}\binom{p+k}{p}\Delta^{k+1}\mu_p\right| < M$$
$$(m = 0, 1, 2, \cdots; k = 0, 1, 2, \cdots),$$

then the limits

$$\lim_{q\to\infty}\binom{q+k}{q}\Delta^k\mu_{q+1} \tag{10}$$

exist ($k = 0, 1, 2, \cdots$) *and, except perhaps for that corresponding to* $k = 0$, *all vanish.*

We note here for future use the following properties of the functions $L_{k,t}\{\mu_n\}$ and $S_{k,t}\{\mu_n\}$.

(11) $L_{k,t}\{\mu_n\}$ is a step function, with jumps of amount

$$\frac{(i+k+1)!}{i!\,k!}(-1)^k\Delta^k\mu_i - \frac{(i+k)!}{(i-1)!\,k!}(-1)^k\Delta^k\mu_{i-1}$$

at the points $i/(i+k)$ ($i = 1, 2, \cdots; k > 0$).

[16] D. V. Widder, op. cit., p. 189.

(12) $S_{k,t}\{\mu_n\}$ is a step function, with jumps of amount

$$(-1)^{k+1}\binom{i+k}{k}\Delta^{k+1}\mu_i$$

at the points $i/(i+k)$ $(i = 0, 1, 2, \cdots)$; $S_{k,0}\{\mu_n\} = -\mu_0$ by definition.

We are now in a position to establish

THEOREM 1. *For every sequence* $\{\mu_n\}$ *such that*

$$\sum_{i=0}^{\infty}\binom{i+k}{k}|\Delta^{k+1}\mu_i| < K \qquad (k = 0, 1, 2, \cdots) \tag{13}$$

it is true that

$$\begin{aligned}\mu_m - \mu_\infty &= \lim_{k\to\infty}\int_0^1 t^m dS_{k,t}\{\mu_n\} \\ &= \lim_{k\to\infty}\sum_{i=0}^{\infty}\binom{i+k}{k}\left(\frac{i}{i+k}\right)^m(-1)^{k+1}\Delta^{k+1}\mu_i.\end{aligned} \tag{14}$$

The series and the integral in (14) are seen to be equal by (12).

This is a special case of a known theorem.[17] Since the proof can be given much more simply in this more restricted case, we give it here.

By Lemma 7, we know that under (13) the limits $\lim_{q\to\infty}\binom{q+k}{k}\Delta^k\mu_{q+1}$ exist, and, except perhaps for the one corresponding to $k = 0$, i.e., μ_∞, are all zero. From this it follows by partial summation that

$$\sum_{i=0}^{\infty}\binom{i+k}{k}(-1)^{k+1}\Delta^{k+1}\mu_i = \mu_0 - \mu_\infty, \tag{15}$$

and hence that

$$\mu_0 - \mu_\infty = \lim_{k\to\infty}\sum_{i=0}^{\infty}\binom{i+k}{k}(-1)^{k+1}\Delta^{k+1}\mu_i.$$

This is the conclusion of the theorem for $m = 0$. We now proceed by induction. Assume the theorem proved for $0, 1, 2, \cdots, m$; we establish it for $m + 1$.

Since $i/(i+k) < 1$, $k > 0$, the series on the right of (14) converges whenever (13) holds. We have

$$\begin{aligned}&\sum_{i=0}^{\infty}\binom{i+k}{k}\left(\frac{i}{i+k}\right)^{m+1}(-1)^{k+1}\Delta^{k+1}\mu_i \\ &\qquad = \sum_{i=0}^{\infty}\binom{i+k}{k}\left(\frac{i+1}{i+k+1}\right)^m(-1)^{k+1}\Delta^{k+1}\mu_{i+1}.\end{aligned}$$

[17] D. V. Widder, op. cit., p. 188.

Now it is seen at once that if (13) is true for the sequence $\{\mu_n\}$, it is true also for the sequence $\{\mu_{n+1}\}$. By the induction hypothesis, then,

$$\lim_{k\to\infty} \sum_{i=0}^{\infty} \binom{i+k}{k} \left(\frac{i}{i+k}\right)^m (-1)^{k+1}\Delta^{k+1}\mu_{i+1} = \mu_{m+1} - \mu_\infty .$$

Now,

$$\left| \sum_{i=0}^{\infty} \binom{i+k}{k} \left[\left(\frac{i+1}{i+k+1}\right)^m - \left(\frac{i}{i+k}\right)^m \right] (-1)^{k+1}\Delta^{k+1}\mu_{i+1} \right|$$

$$\leqq \sum_{i=0}^{\infty} \binom{i+k}{k} \frac{k}{(i+k+1)^m (i+k)^m}$$

$$\sum_{j=1}^{m} [(i+k)(i+1)]^{m-j} [i(i+k+1)]^{j-1} \mid \Delta^{k+1}\mu_{i+1} \mid$$

$$\leqq \frac{m}{k} \sum_{i=0}^{\infty} \binom{i+k}{k} \mid \Delta^{k+1}\mu_{i+1} \mid \leqq \frac{Km}{k} \to 0 \qquad (k \to \infty) .$$

This establishes the induction, and the theorem is proved.

COROLLARY. *For every sequence* $\{\mu_n\}$ *for which* (13) *holds, and for every polynomial* $P_n(x)$,

$$M[P_n(x)] - \mu_\infty P_n(1) = \lim_{k\to\infty} \int_0^1 P_n(t) dS_{k,t}\{\mu_n\} . \tag{16}$$

3. **The deduction of criteria** (A), (B) **and** (C). We first prove

THEOREM 2.[18] *A necessary and sufficient condition for* (1) *to have a solution* $\alpha(t)$ *of bounded variation on* (0, 1) *is that for some* $K > 0$

$$\int_0^1 \mid L_{k,t}\{\mu_n\} \mid dt < K \qquad (k = 0, 1, 2, \cdots) . \tag{A}$$

We establish first the necessity of the condition. Since $L_{0,t}\{\mu_n\} = \mu_0$, $0 \leqq t \leqq 1$, we may suppose henceforward that $k > 0$. Set

$$\nu_n = \int_0^1 t^n dV(t) \qquad (n = 0, 1, 2, \cdots) ,$$

where $V(t)$ is the total variation of $\alpha(u)$ on $0 \leqq u \leqq t$. Then

$$\mid \Delta^k \mu_i \mid = \left| \int_0^1 t^i (1-t)^k d\alpha(t) \right| \leqq \int_0^1 t^i(1-t)^k dV(t) = (-1)^k \Delta^k \nu_i .$$

[18] D. V. Widder, op. cit., p. 194.

By referring to the property (11) of $L_{k,t}\{\mu_n\}$ we see that

$$(17)\qquad \int_0^1 | L_{k,t}\{\mu_n\} | \, dt = \sum_{i=0}^{\infty} \frac{(i+k+1)!}{i!k!} | \Delta^k \mu_i | \left\{ \frac{i+1}{i+k+1} - \frac{i}{i+k} \right\}$$

$$= \sum_{i=0}^{\infty} \binom{i+k-1}{i} | \Delta^k \mu_i |$$

$$\leqq \sum_{i=0}^{\infty} \binom{i+k-1}{i} (-1)^k \Delta^k \nu_i ,$$

provided that this last series is convergent. We saw in the proof of Theorem 1 that

$$\sum_{i=0}^{\infty} \binom{i+k-1}{i} (-1)^k \Delta^k \nu_i = \nu_0 - \nu_\infty$$

provided that

$$(18)\qquad \lim_{q\to\infty} D_{q,k} = \lim_{q\to\infty} (-1)^k \binom{q+k}{k} \Delta^k \nu_{q+1} = 0 \qquad (k = 1, 2, \cdots).$$

We proceed to establish (18). When this has been done, we shall have from (17)

$$\int_0^1 | L_{k,t}\{\mu_n\} | \, dt \leqq \nu_0 - \nu_\infty \qquad (k = 1, 2, \cdots).$$

We have

$$D_{q,k} = \binom{q+k}{k} \int_0^1 t^{q+1}(1-t)^k dV(t) = \int_0^{1-} \binom{q+k}{k} t^{q+1}(1-t)^k dV(t) ,$$

since $(1-t)^k t^{q+1}$ vanishes at $t = 1$. Then,

$$D_{q,k} = \int_0^{1-\delta} \binom{q+k}{k} t^{q+1}(1-t)^k dV(t) + \int_{1-\delta}^{1-} \binom{q+k}{k} t^{q+1}(1-t)^k dV(t)$$

$$\equiv I_1 + I_2 . \qquad (0 < \delta < 1)$$

Now the maximum of $t^{q+1}(1-t)^k$ occurs for $t = (q+1)/(q+k+1)$, and simple computation shows that there is a number $A(k)$ depending only on k such that

$$\binom{q+k}{k} t^{q+1} (1-t)^k \leqq A(k) \qquad (0 \leqq t \leqq 1).$$

Hence

$$I_2 \leqq A(k) \int_{1-\delta}^{1-} dV(t) = A(k)\{ V(1-) - V(1-\delta)\} < \epsilon ,$$

for an arbitrary preassigned $\epsilon > 0$ and for δ chosen sufficiently small. Now we fix δ, and note that

$$\lim_{q\to\infty} \binom{q+k}{k} t^{q+1}(1-t)^k = 0,$$

uniformly for $0 \leqq t \leqq 1-\delta$. We may then take the limit under the integral sign in I_1, so that $\lim\limits_{q\to\infty} I_1 = 0$. Hence

$$\lim_{q\to\infty} D_{q,k} = 0 \qquad (k = 1, 2, \cdots),$$

which completes the proof of the necessity of (A). We now establish the sufficiency.

Supposing (A) satisfied, we have, by the corollary of Theorem 1, for every polynomial $P_n(x)$,

$$(19)\qquad M[P_n(x)] = \mu_\infty P_n(1) + \lim_{k\to\infty} \sum_{i=0}^{\infty} \binom{i+k}{k} P_n\left(\frac{i}{i+k}\right)(-1)^{k+1}\Delta^{k+1}\mu_i,$$

$$|M[P_n(x)]| \leqq |\mu_\infty P_n(1)| + \lim_{k\to\infty} \sum_{i=0}^{\infty} \binom{i+k}{k} \left| P_n\left(\frac{i}{i+k}\right)\right| \cdot |\Delta^{k+1}\mu_i|$$

$$\leqq (|\mu_\infty| + K) \| P_n \|_A = K' \| P_n \|_A .$$

This is condition (7). This establishes the existence of the desired function $\alpha(t)$.

Theorem 3. *A necessary and sufficient condition for* (3) *to have a solution $\varphi(t)$ in $L^{(r)}$, $r > 1$, is that for some $K > 0$,*

$$\text{(B)}\qquad \int_0^1 |L_{k,t}\{\mu_n\}|^r dt < K^r \qquad (k = 0, 1, 2, \cdots).$$

We prove first the necessity of the condition. It is trivial for $k = 0$. Since $\varphi(t)$ is given as a solution of (3), we have, for $k > 0$, $0 \leqq t < 1$,

$$L_{k,t}\{\mu_n\} = \frac{(n+k+1)!}{n!\,k!}(-1)^k\Delta^k\mu_n$$

$$= \frac{(n+k+1)!}{n!\,k!}\int_0^1 u^n(1-u)^k\varphi(u)du, \quad n = \left[\frac{kt}{1-t}\right].$$

Now,

$$(20)\qquad \int_0^1 |\varphi(u)|^r du = \int_0^1 \sum_{p=0}^{\infty} \binom{p+k-1}{p} u^p(1-u)^k |\varphi(u)|^r du,$$

for any given $k > 0$, since

$$\sum_{p=0}^{\infty} \binom{p+k-1}{p} t^p(1-t)^k = 1 \qquad (0 \leqq t < 1);$$

moreover, this series is boundedly convergent, so that we may interchange the order of summation and integration in (20), obtaining

$$\int_0^1 |\varphi(u)|^r du = \sum_{p=0}^{\infty} \int_0^1 \binom{p+k-1}{p} (1-u)^k u^p \, |\varphi(u)|^r du$$

$$= \sum_{p=0}^{\infty} \frac{k}{(p+k)(p+k+1)} \int_0^1 \frac{(p+k+1)!}{p!\,k!} (1-u)^k u^p \, |\varphi(u)|^r du$$

$$\geqq \sum_{p=0}^{\infty} \frac{k}{(p+k)(p+k+1)} \left| \frac{(p+k+1)!}{p!\,k!} \int_0^1 (1-u)^k u^p \varphi(u) dn \right|^r$$

$$= \sum_{p=0}^{\infty} \frac{k}{(p+k)(p+k+1)} \left| \frac{(p+k+1)!}{p!\,k!} \Delta^k u_p \right|^r,$$

by use of the inequality[19]

$$\left(\int_0^1 \lambda(x)\, |\varphi(x)|\, dx\right)^r \leqq \int_0^1 \lambda(x)\, |\varphi(x)|^r dx \qquad \left(\lambda(x) \geqq 0\,, \quad \int_0^1 \lambda(x) dx = 1\right).$$

Here

$$\lambda(t) = \frac{(p+k+1)!}{p!\,k!} (1-t)^k t^p\,.$$

Hence, noting the form of $L_{k,t}\{\mu_n\}$, as given by (11), we have as in (17)

$$K^r = \int_0^1 |\varphi(u)|^r du \geqq \sum_{p=0}^{\infty} \left(\frac{p+1}{p+k+1} - \frac{p}{p+k}\right) | L_{k,p/p+k}\{\mu_n\} |^r$$

$$= \int_0^1 | L_{k,t}\{\mu_n\} |^r \, dt\,.$$

This establishes the necessity of (B). We turn now to the proof of the sufficiency.

We are now assuming

$$(21) \quad \sum_{p=0}^{\infty} \frac{k}{(p+k)(p+k+1)} \left| \frac{(p+k+1)!}{p!\,k!} \Delta^k \mu_p \right|^r < K^r \quad (k = 0, 1, 2, \cdots).$$

We shall show that

$$(22) \qquad \sum_{p=0}^{\infty} \binom{p+k-1}{p} | \Delta^k \mu_p | < K\,,$$

[19] Used in a similar connection by Hausdorff, op. cit., p. 233.

uniformly for $k = 1, 2, \cdots$. To this end, consider

$$\sum_{p=0}^{P} \binom{p+k-1}{p} | \Delta^k \mu_p | = \sum_{p=0}^{P} \left(\frac{(p+k)(p+k+1)}{k}\right)^{1/s}$$

$$\left(\frac{k}{(p+k)(p+k+1)}\right)^{1/s} \binom{p+k-1}{p} | \Delta^k \mu_p |, \qquad \left(\frac{1}{r} + \frac{1}{s} = 1\right)$$

$$\leqq \left(\sum_{p=0}^{P} \left(\frac{(p+k)(p+k+1)}{k}\right)^{r-1} \left| \binom{p+k-1}{p} \Delta^k \mu_p \right|^r \right)^{1/r}$$

$$\left(\sum_{p=0}^{P} \frac{k}{(p+k)(p+k+1)}\right)^{1/s}$$

$$\leqq \left(\sum_{p=0}^{\infty} \frac{k}{(p+k)(p+k+1)} \left| \frac{(p+k+1)!}{p!\,k!} \Delta^k \mu_p \right|^r \right)^{1/r}$$

$$\left(\sum_{p=0}^{\infty} \frac{k}{(p+k)(p+k+1)}\right)^{1/s}$$

$$\leqq K,$$

by (21), since $\sum_{p=0}^{\infty} k/(p+k)(p+k+1) = 1, k > 0$.

This holds for any $P > 0$; hence the series in (22) converges, and relation (22) is valid.

Now (22) is (13) with k replaced by $k - 1$, and the corollary of Theorem 1 applies; that is, for any polynomial $P_n(t)$

$$(23) \quad |M[P_n] - \mu_\infty P_n(1)| = \lim_{k\to\infty} \left| \sum_{p=0}^{\infty} \binom{p+k-1}{p} P_n\left(\frac{p}{p+k-1}\right) (-1)^k \Delta^k \mu_p \right|,$$

where it is to be noted that because of (22) the series on the right is convergent. Now,

$$\left| \sum_{p=0}^{\infty} \binom{p+k-1}{p} P_n\left(\frac{p}{p+k-1}\right) (-1)^k \Delta^k \mu_p \right|$$

$$= \left| \sum_{p=0}^{\infty} \binom{p+k-1}{p} P_n\left(\frac{p}{p+k-1}\right) \left(\frac{k}{(p+k)(p+k+1)}\right)^{1/s} \right.$$

$$\left. \left(\frac{(p+k)(p+k+1)}{k}\right)^{1/s} (-1)^k \Delta^k \mu_p \right|$$

$$\leqq \left(\sum_{p=0}^{\infty}\left(\frac{(p+k)(p+k+1)}{k}\right)^{r-1}\left|\binom{p+k-1}{p}\Delta^k\mu_p\right|^r\right)^{1/r}\cdot$$

$$\left(\sum_{p=0}^{\infty}\left|P_n\left(\frac{p}{p+k-1}\right)\right|^s\frac{k}{(p+k)(p+k+1)}\right)^{1/s}$$

$$\leqq K\cdot\left\{\sum_{p=0}^{\infty}\left|P_n\left(\frac{p}{p+k-1}\right)\right|^s\left(\frac{p+1}{p+k+1}-\frac{p}{p+k}\right)\right\}^{1/s}$$

$$\equiv K(G_k)^{1/s} \qquad \left(\frac{1}{r}+\frac{1}{s}=1\right).$$

It is to be noted that the series denoted by G_k is actually convergent, because $|P_n(t)|$ is bounded; furthermore, it is easily seen that

$$\lim_{k\to\infty} G_k = \int_0^1 |P_n(t)|^s\,dt = (\|P_n\|_B)^s.$$

Hence, returning to (23) we have

$$|M[P_n(t)] - \mu_\infty P_n(1)| \leqq K\|P_n\|_B,$$

which is the condition (8), sufficient to ensure the existence of a function $\varphi(t)$ of $L^{(r)}$ such that

$$\mu_n - \mu_\infty = \int_0^1 t^n\varphi(t)dt \qquad (n = 0, 1, 2, \cdots).$$

The theorem is thus established.

THEOREM 4.[20] *A necessary and sufficient condition for* (3) *to have a solution* $\varphi(t)$ *bounded almost everywhere on* $(0, 1)$ *is that for some* $K > 0$,

(C) $$|L_{k,t}\{\mu_n\}| < K \qquad (k = 0, 1, 2, \cdots; 0 \leqq t \leqq 1).$$

We establish first the necessity of (C).

$$|L_{k,t}\{\mu_n\}| = \frac{(n+k+1)!}{n!\,k!}|\Delta^k\mu_n| \qquad \left(n = \left[\frac{kt}{1-t}\right], 0 \leqq t < 1\right)$$

$$= \frac{(n+k+1)!}{n!\,k!}\left|\int_0^1 (1-u)^k u^n\varphi(u)\,du\right|$$

$$\leqq \frac{(n+k+1)!}{n!\,k!}\cdot \operatorname*{true\,max}_{0\leqq t\leqq 1}|\varphi(u)|\int_0^1 (1-u)^k u^n du^{20a}$$

$$= \operatorname*{true\,max}_{0\leqq t\leqq 1}|\varphi(u)| \leqq K \qquad (k = 1, 2, \cdots).$$

[20] F. Hausdorff, op. cit., p. 237.

[20a] True max $|f(x)| = M$ if $|f(x)| \leqq M$ almost everywhere, and if for every $\epsilon > 0$, $|f(x)| \geqq M - \epsilon$ on a set of positive measure.

For $k = 0, n = 0$, and $| L_{0,t}\{\mu_n\} | = | \mu_0 | \leqq K$; for $t = 1$, (C) holds by continuity.

We turn now to the sufficiency. If (C) holds, for any $P > 0$

$$\sum_{p=0}^{P} \binom{p+k-1}{p} | \Delta^k \mu_p | = \sum_{p=0}^{\infty} \frac{k}{(p+k)(p+k+1)} \frac{(p+k+1)!}{p!\,k!} | \Delta^k \mu_p |$$

$$\leqq K \sum_{p=0}^{\infty} \frac{k}{(p+k)(p+k+1)} = K \qquad (k = 1, 2, \cdots).$$

Hence

$$\sum_{p=0}^{\infty} \binom{p+k-1}{p} | \Delta^k \mu_p | \leqq K \qquad (k = 1, 2, \cdots). \tag{24}$$

(24) is (13) with k replaced by $k - 1$; we may again apply the corollary of Theorem 1, so that for any polynomial $P_n(t)$ we have, applying (C),

$$| M[P_n(t)] - \mu_\infty P_n(1) | = \lim_{k\to\infty} \left| \sum_{p=0}^{\infty} \binom{p+k-1}{p} P_n\left(\frac{p}{p+k}\right) (-1)^k \Delta^k \mu_p \right|$$

$$\leqq K \lim_{k\to\infty} \sum_{p=0}^{\infty} \left| P_n\left(\frac{p}{p+k}\right) \right| \frac{k}{(p+k)(p+k+1)}$$

$$= K \, \| P_n \|_C ,$$

as in the last theorem. This is the condition (9), which ensures the existence of a solution $\varphi(t)$ of (3), bounded almost everywhere.

4. **The criteria** (D) **and** (E). The remaining criteria cannot be directly obtained from theorems on linear operations, since there are no relevant theorems available. The cases to be considered are all sub-cases of those already treated, and the criteria in each case result from an application of the more general criteria already found, together with an examination of the specific properties of the special classes of functions involved.

We shall need an additional inversion formula for moment sequences of the form (1). We state it as

LEMMA 8. *If*

$$\mu_n = \int_0^1 t^n d\alpha(t) \qquad (n = 0, 1, 2, \cdots),$$

where $\alpha(t)$ is a function of bounded variation, $0 \leqq t \leqq 1$, and $\alpha(1) = 0$, then

$$\frac{\alpha(t+) + \alpha(t-)}{2} = -\mu_0 + \lim_{k\to\infty} \int_0^t L_{k,u}\{\mu_n\}\, du \qquad (0 < t < 1). \tag{25}$$

Proof. By Lemma 4,

$$\frac{\alpha(t+)+\alpha(t-)}{2} = \lim_{k\to\infty} S_{k,t}\{\mu_n\} \qquad (0 < t < 1).$$

$$\begin{aligned} S_{k,t}\{\mu_n\} &= -\mu_\infty - \sum_{i=n+1}^{\infty} \frac{(i+k)!}{i!\,k!}(-1)^{k+1}\Delta^{k+1}\mu_i, \qquad n = \left[\frac{kt}{1-t}\right] \\ &= -\mu_\infty + \sum_{i=0}^{n}\binom{i+k}{k}(-1)^{k+1}\Delta^{k+1}\mu_i - \sum_{i=0}^{\infty}\binom{i+k}{k}(-1)^{k+1}\Delta^{k+1}\mu_i \\ &= -\mu_\infty + \sum_{i=0}^{n}\binom{i+k}{k}(-1)^{k+1}\Delta^{k+1}\mu_i - \mu_0 + \mu_\infty, \end{aligned}$$

since the relation (18) established in Theorem 2 allows us to conclude that the limits (10) of Lemma 7 vanish, and hence that the relation (15) of Theorem 1 is valid. We thus have

$$\begin{aligned} S_{k,t}\{\mu_n\} &= -\mu_0 + \sum_{i=0}^{n}\binom{i+k}{k}(-1)^{k+1}\Delta^{k+1}\mu_i \\ &= -\mu_0 + \int_0^{\frac{n+1}{n+k+2}} L_{k+1,u}\{\mu_n\}\,du, \qquad n = \left[\frac{kt}{1-t}\right], \end{aligned}$$

as we see by property (11) of $L_{k,t}\{\mu_n\}$. Our lemma will therefore be established if we show that for each fixed t, $0 < t < 1$,

$$\begin{aligned} \lim_{k\to\infty} I_k(t) = \lim_{k\to\infty}\left| -\mu_0 + \int_0^t L_{k+1,u}\{\mu_n\}\,du - S_{k,t}\{\mu_n\}\right| \\ = \lim_{k\to\infty}\left|\int_{\frac{n+1}{n+k+2}}^{t} L_{k+1,u}\{\mu_n\}\,du\right| = 0. \end{aligned}$$

We have

$$I_k(t) \leqq \int_{\frac{n+1}{n+k+2}}^{t} |L_{k+1,u}\{\mu_n\}|\,du. \tag{26}$$

But we also have

$$L_{k+1,u}\{\mu_n\} = \frac{(n'+k+2)!}{n'!\,(k+1)!}\int_0^1 t^{n'}(1-t)^{k+1}\,d\alpha(t), \qquad n' = \left[\frac{(k+1)u}{1-u}\right].$$

Since the maximum of $t^{n'}(1-t)^{k+1}$ occurs for $t = n'/(n'+k+1)$,

$$|L_{k+1,u}\{\mu_n\}| \leqq A\,\frac{(n'+k+2)!}{n'!\,(k+1)!}\left(\frac{n'}{n'+k+1}\right)^{n'}\left(\frac{k+1}{n'+k+1}\right)^{k+1}, \tag{27}$$

where A is the total variation of $\alpha(t)$ on $0 \leqq t \leqq 1$. Since u is restricted to

lie between $\dfrac{n+1}{n+k+2}$ and t, an upper bound for the right hand side of (27) may be easily obtained by Stirling's formula, and we find for the values u under consideration,

$$| L_{k+1,u}\{\mu_n\} | = O(\sqrt{k}) \qquad (k \to \infty) .$$

Hence, by (26),

$$I_k(t) = O\left\{\sqrt{k}\left(t - \frac{n+1}{n+k+2}\right)\right\} \qquad (k \to \infty) . \tag{28}$$

It is easily seen that

$$\left| t - \frac{n+1}{n+k+2} \right| = O\left(\frac{1}{k}\right) \qquad (k \to \infty) ,$$

and hence by (28) that $I_k(t)$ has the limit zero as $k \to \infty$. This establishes the lemma.

We are now in a position to establish

THEOREM 5. *A necessary and sufficient condition that* (3) *have a solution* $\varphi(t)$ *integrable over* (0, 1) *is that*

$$\lim_{k,k' \to \infty} \int_0^1 | L_{k,t}\{\mu_n\} - L_{k',t}\{\mu_n\} | dt = 0 . \tag{D}$$

We prove first the necessity. We are given that

$$\mu_n - \mu_\infty = \int_0^1 t^n \varphi(t) dt \qquad (n = 0, 1, 2, \cdots) .$$

The series

$$\sum_{i=0}^{\infty} \frac{(i+k-1)!}{i!\,(k-1)!} y^i (1-y)^k \qquad (= 1, 0 \leqq y < 1)$$

has its partial sums uniformly bounded with respect to y and k, as we noted before. Hence, for any integrable function $\omega(y)$,

$$\begin{aligned} \int_0^1 | \omega(y) | dy &= \int_0^1 | \omega(y) | \sum_{i=0}^{\infty} \frac{(i+k-1)!}{i!\,(k-1)!} y^i(1-y)^k \, dy \\ &= \sum_{i=0}^{\infty} \binom{i+k-1}{i} \int_0^1 | \omega(y) | y^i(1-y)^k \, dy \\ &\geqq \sum_{i=0}^{\infty} \binom{i+k-1}{i} \left| \int_0^1 \omega(y) y^i (1-y)^k dy \right| , \end{aligned}$$

$$\int_0^1 | \omega(y) | dy \geqq \int_0^1 | L_{k,t}\{\pi_n\} | dt , \tag{29}$$

where $\pi_n = \int_0^1 t^n\omega(t)dt$. Hence, taking $\omega(y) = \varphi(y)$,

$$\int_0^1 |\varphi(y)| \, dy \geqq \int_0^1 | L_{k,t}\{\mu_n\} | \, dt. \tag{30}$$

Given $\epsilon > 0$, we can find a continuous $\psi(y)$ such that

$$\int_0^1 |\varphi(y) - \psi(y)| \, dy < \epsilon. \tag{31}$$

Also, by Lemma 6, if we set

$$\nu_n = \int_0^1 t^n\psi(t)dt,$$

we have $\lim_{k\to\infty} L_{k,t}\{\nu_n\} = \psi(t)$, uniformly on $0 \leqq t \leqq 1$, and, *à fortiori*,

$$\lim_{k\to\infty} \int_0^1 | L_{k,t}\{\nu_n\} - \psi(t) | \, dt = 0. \tag{32}$$

Furthermore, the inequality (29), if applied to the function $\omega(y) = \{\varphi(y) - \psi(y)\}$, shows that

$$\begin{aligned}\int_0^1 | L_{k,t}\{\mu_n\} - L_{k,t}\{\nu_n\} | \, dt &= \int_0^1 | L_{k,t}\{\mu_n - \nu_n\} | \, dt \\ &\leqq \int_0^1 | \varphi(y) - \psi(y) | \, dy.\end{aligned} \tag{33}$$

We then have

$$\begin{aligned}\int_0^1 |\varphi(t) - L_{k,t}\{\mu_n\} | \, dt \leqq \int_0^1 | \varphi(t) - \psi(t) | \, dt &+ \int_0^1 | \psi(t) - L_{k,t}\{\nu_n\} | \, dt \\ &+ \int_0^1 | L_{k,t}\{\mu_n\} - L_{k,t}\{\nu_n\} | \, dt,\end{aligned}$$

and for k sufficiently large, by (31), (32), and (33),

$$\int_0^1 | \varphi(t) - L_{k,t}\{\mu_n\} | \, dt \leqq 3\epsilon.$$

Hence

$$\lim_{k\to\infty} \int_0^1 | \varphi(t) - L_{k,t}\{\mu_n\} | \, dt = 0, \tag{34}$$

and from (34) the necessity of (D) follows immediately.

We now establish the sufficiency of (D). By hypothesis,

$$\lim_{k,k'\to\infty} \int_0^1 | L_{k,t}\{\mu_n\} - L_{k',t}\{\mu_n\} | \, dt = 0.$$

Let $\alpha_k(t) = \int_0^t | L_{k,u}\{\mu_n\} | \, du$. Then, for any t, $0 \leqq t \leqq 1$,

$$| \alpha_k(t) - \alpha_{k'}(t) | \leqq \int_0^t || L_{k,u}\{\mu_n\} | - | L_{k',u}\{\mu_n\} || \, du$$
$$\leqq \int_0^1 | L_{k,u}\{\mu_n\} - L_{k',u}\{\mu_n\} | \, du .$$

Thus the sequence $\{\alpha_k(t)\}$ converges uniformly on $0 \leqq t \leqq 1$. Hence, *à fortiori*,

$$| \alpha_k(t) | < K, \; K > 0 ,$$

uniformly in t and k. If we now express $\alpha_k(1)$ as a series (using (11)), we find

$$\sum_{i=0}^{\infty} \binom{i + k - 1}{i} | \Delta^k \mu_i | < K \qquad (k = 1, 2, \cdots) \tag{35}$$

Now set $\beta_k(t) = \int_0^t L_{k,u}\{\mu_n\} du$. Then it is clear that the sequence $\{\beta_k(t)\}$ converges uniformly on $0 \leqq t \leqq 1$ to a continuous function $\beta(t)$. We shall show that $\beta(t)$ is absolutely continuous.[21] To this end, let $\epsilon > 0$ and a set E of non-overlapping intervals (t_i, t_i') be given.

$$\sum_i | \beta_k(t_i') - \beta_k(t_i) - \beta_{k'}(t_i') + \beta_{k'}(t_i) | = \sum_i \left| \int_{t_i}^{t_i'} (L_{k,t}\{\mu_n\} - L_{k',t}\{\mu_n\}) \, dt \right|$$
$$\leqq \int_0^1 | L_{k,t}\{\mu_n\} - L_{k',t}\{\mu_n\} | \, dt < \epsilon ,$$

for k and k' sufficiently large. Let k be chosen sufficiently large and then fixed; let $k' \to \infty$. We obtain

$$\epsilon \geqq \sum_i | \beta_k(t_i') - \beta_k(t_i) - \beta(t_i') + \beta(t_i) |$$
$$\geqq | \sum_i | \beta_k(t_i') - \beta_k(t_i) | - \sum_i | \beta(t_i') - \beta(t_i) | | .$$

Here k is fixed; since $\beta_k(t)$ is absolutely continuous, we deduce that, if the measure of E is sufficiently small,

$$\sum_i | \beta(t_i') - \beta(t_i) | < 2\epsilon ,$$

so that $\beta(t)$ is absolutely continuous.

On the other hand, (35) is equivalent to (A), and by Theorem 2 implies that

$$\mu_n = \int_0^1 t^n d\gamma(t) \qquad (n = 0, 1, 2, \cdots) , \tag{36}$$

[21] The proof of this fact follows in essentials the outline of the proof of a similar result given by Hausdorff, op. cit., p. 247.

where $\gamma(t)$ is a function of bounded variation. By Lemma 8,

$$\gamma(t) = -\mu_0 + \lim_{k\to\infty} \int_0^t L_{k,u}\{\mu_n\}\, du, \qquad (0 < t < 1).$$

But

$$\beta(t) = \lim_{k\to\infty} \int_0^t L_{k,u}\{\mu_n\}\, du$$

by definition. Hence

$$\beta(t) - \gamma(t) = \mu_0, \qquad 0 < t < 1.$$

Since $\beta(t)$ is absolutely continuous, $\gamma(t)$ is absolutely continuous for $0 < t < 1$, and since we may modify $\gamma(t)$ by an additive constant to make $\gamma(0) = \gamma(0+)$ without affecting (36),

$$\mu_n = \int_{1-}^1 t^n d\gamma(t) + \int_0^1 t^n \varphi(t) dt,$$

where $\varphi(t) = \gamma'(t)$ for $0 < t < 1$, wherever the derivative exists, and is suitably defined at the remaining points of (0, 1) and at the end points. That is,

$$\mu_n = \gamma(1) - \gamma(1-) + \int_0^1 t^n \varphi(t) dt.$$

Let $n \to \infty$; the integral approaches zero, so that $\gamma(1) - \gamma(1-) = \mu_\infty$, and

$$\mu_n - \mu_\infty = \int_0^1 t^n \varphi(t) dt \qquad (n = 0, 1, 2, \cdots).$$

The theorem is thus fully established.

THEOREM 6. *A necessary and sufficient condition that* (3) *have a solution* $\varphi(t)$ *of bounded variation,* $0 \leqq t \leqq 1$, *is that for some* $K > 0$,

$$\text{(E)} \qquad \int_0^1 |\, dL_{k,t}\{\mu_n\}\,| < K \qquad (k = 0, 1, 2, \cdots),$$

provided that we redefine $L_{k,0}\{\mu_n\} = 0$.

Necessity. Given

$$\mu_n - \mu_\infty = \int_0^1 t^n \varphi(t) dt \qquad (n = 0, 1, 2, \cdots)$$

assume $\varphi(1) = 0$; this produces no change in the μ_n. Consider

$$\text{(37)} \qquad \nu_n = \int_0^1 t^n d\varphi(t) \qquad (n = 1\ 2, \cdots)$$

$$= -n \int_0^1 t^{n-1} \varphi(t) dt.$$

Then

$$\mu_n - \mu_\infty = -\frac{\nu_{n+1}}{n+1}.$$

It may be established[22] that

(38) $\quad L_{k,t}\{\mu_n\} = L_{k,t}\{\mu_n - \mu_\infty\} = S_{k,t}\{\nu_n\}, \qquad (k = 1, 2, \cdots; 0 < t < 1).$

By Theorem 2, since (37) has a solution of bounded variation,

(39) $$K \geqq \int_0^1 | L_{k+1,t}\{\nu_n\} | \, dt \qquad (k = 0, 1, \cdots).$$

But we have

$$\int_0^1 | L_{k+1,t}\{\nu_n\} | \, dt = \sum_{i=0}^\infty \binom{i+k}{k} | \Delta^{k+1} \nu_i | = \int_0^1 | dS_{k,t}\{\nu_n\} |$$
$$(k = 1, 2, \cdots),$$

by using (11) and (12); cf. (14) and (17). Referring to (38) and (39), we have

$$K \geqq \int_{0+}^1 | dL_{k,t}\{\mu_n\} | \qquad (k = 1, 2, \cdots),$$

and hence, since we are taking $L_{k,0}\{\mu_n\} = 0$,

$$K + \max_{0 \leqq t \leqq 1} | \varphi(t) | \geqq \int_0^1 | dL_{k,t}\{\mu_n\} | \qquad (k = 1, 2, \cdots).$$

For $k = 0$, $\int_0^1 | dL_{k,t}\{\mu_n\} | = | \mu_0 | \leqq \max_{0 \leqq t \leqq 1} | \varphi(t) |$.

Sufficiency. Given

$$\int_0^1 | dL_{k,t}\{\mu_n\} | \leqq K \qquad (k = 0, 1, 2, \cdots).$$

Then

$$| L_{k,t} \{\mu_n\} | = \left| \int_0^t dL_{k,u}\{\mu_n\} \right| \leqq \int_0^1 | dL_{k,t}\{\mu_n\} | \leqq K.$$

Hence (C) holds, and

$$\mu_n - \mu_\infty = \int_0^1 t^n \varphi(t) dt \qquad (n = 0, 1, \cdots),$$

where $\varphi(t)$ is bounded almost everywhere. By the inversion formula of Lemma 5, $\varphi(t) = \lim_{k \to \infty} L_{k,t}\{\mu_n\}$ for almost all t on $(0, 1)$. But the $L_{k,t}\{\mu_n\}$ have uniformly bounded variation on $0 \leqq t \leqq 1$; hence $\varphi(t)$, if suitably redefined on a

[22] D. V. Widder, op. cit., p. 180.

set of measure zero, will have bounded variation also. This completes the proof.

5. **Further criteria.** We establish now the condition (F).

THEOREM 7. *A necessary and sufficient condition that* (3) *have a solution* $\varphi(t)$ *continuous for* $0 \leqq t \leqq 1$ *is that*

(F) $\{L_{k,t}\{\mu_n\}\}$ *be a uniformly convergent sequence,* $0 \leqq t \leqq 1$.

The sufficiency of this condition would of course be entirely trivial if the functions $L_{k,t}\{\mu_n\}$ were continuous; its chief interest, and the chief difficulty in proving it, come from the fact that these functions are actually step functions.

The necessity of the condition is established in Lemma 6.

Sufficiency. Let $\varphi(t) = \lim_{k\to\infty} L_{k,t}\{\mu_n\}$; the functions $L_{k,t}\{\mu_n\}$ are uniformly bounded for $0 \leqq t \leqq 1$, and hence for some $K > 0$, $|\varphi(t)| < K$. Condition (C) is fulfilled, and therefore there exists a function $\psi(t)$, bounded almost everywhere, such that

$$\mu_n - \mu_\infty = \int_0^1 t^n\psi(t)dt \qquad (n = 0, 1, 2, \cdots).$$

Since by Lemma 5

$$\psi(t) = \lim_{k\to\infty} L_{k,t}\{\mu_n\}$$

for almost all t, $0 \leqq t \leqq 1$, it follows that $\psi(t) = \varphi(t)$ almost everywhere on $(0, 1)$ and

$$\mu_n - \mu_\infty = \int_0^1 t^n\varphi(t)dt \qquad (n = 0, 1, 2, \cdots).$$

It remains to show that $\varphi(t)$ is continuous, $0 \leqq t \leqq 1$. We consider first the open interval. Given any fixed t, $0 < t < 1$, and given $\epsilon > 0$, take k_0 so that $2AK(1-t)/(k_0t)^{1/2} < \epsilon/3$, where A is an absolute constant whose value will be specified later. Then choose $k > k_0$ so large that

$$|\varphi(t) - L_{k,t}\{\mu_n\}| < \epsilon/3$$

for any t in $(0, 1)$ (this is possible because $L_{k,t}\{\mu_n\}$ converges uniformly to $\varphi(t)$), and fix k. Determine $\delta > 0$ so small that $k\delta < 1/2$. Then for any t' such that $|t/(1-t) - t'/(1-t')| < \delta$ we have, if $n = \left[\frac{kt}{1-t}\right]$, $n' = \left[\frac{kt'}{1-t'}\right]$,

$$-\frac{3}{2} \leqq \frac{kt}{1-t} - \frac{kt'}{1-t'} - 1 \leqq n - n' \leqq \frac{kt}{1-t} - \frac{kt'}{1-t'} + 1 \leqq \frac{3}{2}.$$

Thus n and n', being integers, differ by at most unity. If $n = n'$,

$$L_{k,t}\{\mu_n\} - L_{k,t'}\{\mu_n\} = 0.$$

If $n \neq n'$, suppose for definiteness $n' = n + 1$. Then

$$L_{k,t'}\{\mu_n\} - L_{k,t}\{\mu_n\} = \frac{(n+k+2)!}{(n+1)!\,k!}\int_0^1 u^{n+1}(1-u)^k\varphi(u)du$$

$$- \frac{(n+k+1)!}{n!\,k!}\int_0^1 u^n(1-u)^k\varphi(u)du$$

$$= \frac{(n+k+1)!}{n!\,k!}\int_0^1 \left\{\frac{n+k+2}{n+1}u - 1\right\} u^n(1-u)^k\varphi(u)du$$

$$= -\frac{(n+k+1)!}{(n+1)!\,k!}\int_0^1 \frac{d}{du}\{u^{n+1}(1-u)^{k+1}\}\,\varphi(u)du.$$

Hence we have

$$D_k(t) \equiv |\,L_{k,t}\{\mu_n\} - L_{k,t'}\{\mu_n\}\,|$$

$$\leqq K\frac{(n+k+1)!}{(n+1)!\,k!}\int_0^1 \left|\frac{d}{du}\{u^{n+1}(1-u)^{k+1}\}\right| du.$$

Since $u^{n+1}(1-u)^{k+1}$ has a single maximum, at $u = (n+1)/(n+k+2)$, and vanishes at the end points of the interval, the right hand member of this inequality is equal to

$$2K\binom{n+k+1}{k}\frac{(n+1)^{n+1}\,(k+1)^{k+1}}{(n+k+2)^{n+k+2}}.$$

Applying Stirling's formula, we find

$$D_k(t) \leqq AK\sqrt{\frac{k+1}{(n+k+2)(n+1)}},$$

where A is a constant independent of k and t. Since $n = \left[\dfrac{kt}{1-t}\right]$,

$$D_k(t) \leqq AK\sqrt{\frac{k+1}{\dfrac{k}{1-t}\cdot\dfrac{kt}{1-t}}}$$

$$= AK\frac{1-t}{\sqrt{t}}\frac{1}{\sqrt{k}}\sqrt{\frac{k+1}{k}}$$

$$< 2AK\frac{1-t}{\sqrt{t}}\frac{1}{\sqrt{k}} < \epsilon/3.$$

But then

$$|\,\varphi(t) - \varphi(t')\,| \leqq |\,\varphi(t) - L_{k,t}\{\mu_n\}\,| + |\,\varphi(t') - L_{k,t'}\{\mu_n\}\,| + D_k(t) \leqq \epsilon.$$

This establishes the continuity of $\varphi(t)$ at the point t, which was arbitrary in the open interval; a simple argument will now show that, since the sequence $L_{k,t}\{\mu_n\}$ is uniformly convergent in the closed interval, and since each $L_{k,t}\{\mu_n\}$ is con-

tinuous at the two end points, $\varphi(t)$ is continuous at the end points also. This completes the proof.

We can obtain as a corollary of Theorem 7,

THEOREM 8. *A necessary and sufficient condition that the moment problem*

$$\mu_n = \int_0^1 t^n d\alpha(t) \qquad (n = 1, 2, \cdots),$$

with $\lim_{n\to\infty} \mu_n/n = 0$, *have a solution* $\alpha(t)$ *which is continuous* (*not necessarily of bounded variation*) *for* $0 < t < 1$, *with* $\alpha(1) = 0$, *is that*

$$\text{(G)} \qquad L_{k,t}\left\{-\frac{\mu_{n+1}}{n+1}\right\} \text{ be a uniformly convergent sequence on } 0 \leqq t \leqq 1 .$$

This may be simply proved by the use of Theorem 7 and integration by parts. In the same way, by using Theorem 6 one may prove

THEOREM 9. *A necessary and sufficient condition that the moment problem* (2), *with* $\lim_{n\to\infty} n\mu_{n-1} = 0$, *have a solution* $\varphi(t)$ *which is an integral, with* $\varphi(1) = 0$, *is that if we set* $\nu_n = -n\mu_{n-1}$ $(n = 1, 2, \cdots)$ *and make a suitable definition of* ν_0,

$$\text{(H)} \qquad \lim_{k,k'\to\infty} \int_0^1 | L_{k,t}\{\nu_n\} - L_{k',t}\{\nu_n\} | \, dt = 0 ;$$

that is, the sequence $L_{k,t}\{\nu_n\}$ *converge in the mean* (*of order* 1).

6. **Equivalence relations.** We shall now show that the condition (A) and the necessary and sufficient condition of Hausdorff,[23]

$$\text{(A')} \qquad \sum_{i=0}^{p} \binom{p}{i} | \Delta^{p-i}\mu_i | < K \qquad (p = 0, 1, 2, \cdots)$$

are equivalent, in the sense that either is obtainable from the other by algebraic methods, without appeal to the moment problem itself.[24]

THEOREM 10a. *Condition* (A) *implies condition* (A').

We shall consider $p > 0$, since for $p = 0$, (A') is trivial. We show first that

$$(40) \qquad \sum_{n=i}^{\infty} \binom{n-1}{i-1} (-1)^p \Delta^p \mu_n = \begin{cases} (-1)^{p-i}\Delta^{p-i}\mu_i & (p = 1, 2, \cdots; i = 1, 2, \cdots, p-1) \\ \mu_p - \mu_\infty & (i = p) . \end{cases}$$

We noted in the proof of Theorem 1 (see (15)) that for any sequence $\{\nu_n\}$ for which ν_∞ exists and

$$(41) \qquad \lim_{q\to\infty} \binom{q+k}{k} \Delta^k \nu_{q+1} = 0 \qquad (k = 1, 2, \cdots)$$

[23] F. Hausdorff, op. cit., p. 232.

[24] I am indebted to Professor Widder for an outline of the proof of this result.

we have

$$(42)\qquad \sum_{n=0}^{\infty} \binom{n+k-1}{k-1} (-1)^k \Delta^k \nu_n = \nu_0 - \nu_\infty .$$

In the series in (40), set $n - i = j$. The series becomes

$$(43)\qquad \sum_{j=0}^{\infty} \binom{i+j-1}{i-1} (-1)^p \Delta^p \mu_{i+j} .$$

Set

$$\nu_j = (-1)^{p-i} \Delta^{p-i} \mu_{i+j} .$$

Then by (42), the series (43) gives

$$(44)\qquad \sum_{j=0}^{\infty} \binom{i+j-1}{i-1} (-1)^p \Delta^p \mu_{i+j} = \sum_{j=0}^{\infty} \binom{i+j-1}{i-1} (-1)^i \Delta^i \nu_j$$

$$= \begin{cases} (-1)^{p-i} \Delta^{p-i} \mu_i & (i = 1, 2, \cdots, p-1), \\ \mu_p - \mu_\infty & (i = p), \end{cases}$$

which is (40), provided that

$$(45)\qquad \lim_{q\to\infty} \binom{q+k}{k} \Delta^{k+p-i} \mu_{i+q+1} = 0 \qquad (k = 1, 2, \cdots; 0 < i < p).$$

For $k = 0$, $i < p$, (45) is the relation

$$\lim_{q\to\infty} \Delta^{p-i} \mu_{i+q+1} = 0 ,$$

which a simple induction shows to be true; for $k = 0$, $i = p$, the limit on the left of (45) is μ_∞, as it should be.

We now establish (45). By hypothesis,

$$(\mathrm{A})\qquad \sum_{i=0}^{\infty} \binom{i+k}{k} | \Delta^{k+1} \mu_i | < K \qquad (k = 0, 1, 2, \cdots).$$

The series is convergent; its general term must approach zero as i becomes infinite, for each k. In particular,

$$(46)\qquad \lim_{q\to\infty} \binom{k+p+q}{i+q+1} \Delta^{k+p-i} \mu_{i+q+1} = 0 \qquad (k + p - i \geqq 1).$$

(45) will therefore be established if we show that for fixed k and p, for fixed i, $0 < i < p$, and for some $M > 0$,

$$(47)\qquad \frac{(q+k)!}{q!\,k!} \frac{(i+q+1)!\,(k+p-i-1)!}{(k+p+q)!} < M \qquad (q = 1, 2, \cdots).$$

(47) is equivalent to

$$(48) \qquad \frac{(q+k)(q+k-1)\cdots(q+1)}{(k+q+p)(k+p+q-1)\cdots(i+q+2)} < M \frac{k!}{(k+p-i-1)!} \qquad (q = 1, 2, \cdots),$$

which is certainly true because the numerator in (48) has k factors, the denominator has k factors at least for $i < p$, and each factor of the numerator is less than the corresponding factor in the denominator. We may therefore infer the validity of (45), and hence of (40).

We next show that the series (40) is absolutely convergent. We use the form (43), and compare it with

$$\sum_{m=i}^{\infty} \binom{m+p-1}{p-1} | \Delta^p \mu_m | < K \qquad (p = 1, 2, \cdots),$$

which is an immediate consequence of (A); or, setting $m - i = j$, with

$$(49) \qquad \sum_{j=0}^{\infty} \binom{i+j+p-1}{p-1} | \Delta^p \mu_{i+j} | < K \qquad (p = 1, 2, \cdots).$$

Referring to (43), we see that we wish to establish the convergence of

$$(50) \qquad \sum_{j=0}^{\infty} \binom{i+j-1}{i-1} | \Delta^p \mu_{i+j} | \qquad (p = 1, 2, \cdots; i = 1, 2, \cdots, p).$$

To do this, we need only show that for some $A > 0$, and for each pair of values (p, i) considered,

$$\frac{(i+j-1)!}{(i-1)!\,j!} \frac{(p-1)!\,(i+j)!}{(i+j+p-1)!} \leqq A \quad (j = 0, 1, 2, \cdots),$$

or,

$$\frac{(i+j)(i+j-1)\cdots(j+1)\cdot(p-1)(p-2)\cdots(i)}{(i+j+p-1)(i+j+p-2)\cdots(i+j)} \leqq A,$$

which is clearly the case. This establishes the absolute convergence of (40).

We now have, from (40),

$$(51) \begin{cases} | \Delta^{p-i} \mu_i | \leqq \displaystyle\sum_{n=i}^{\infty} \binom{n-1}{i-1} | \Delta^p \mu_n | & (p = 1, 2, \cdots; i = 1, 2, \cdots, p-1); \\ | \Delta^{p-i} \mu_i | \leqq | \mu_\infty | + \displaystyle\sum_{n=p}^{\infty} \binom{n-1}{p-1} | \Delta^p \mu_n | & (i = p); \\ | \Delta^{p-i} \mu_i | = | \Delta^p \mu_0 | & (i = 0). \end{cases}$$

Then,

$$(52)\quad \sum_{i=0}^{p}\binom{p}{i}\,|\,\Delta^{p-i}\mu_i\,| \leqq |\,\mu_\infty\,| + |\,\Delta^p\mu_0\,| + \sum_{i=1}^{p}\binom{p}{i}\sum_{n=i}^{\infty}\binom{n-1}{i-1}\,|\,\Delta^p\mu_n\,|$$

$$= |\,\mu_\infty\,| + |\,\Delta^p\mu_0\,| + \sum_{n=1}^{p}\,|\,\Delta^p\mu_n\,|\sum_{i=1}^{n}\binom{p}{i}\binom{n-1}{i-1}$$

$$+ \sum_{n=p+1}^{\infty}\,|\,\Delta^p\mu_n\,|\sum_{i=1}^{p}\binom{p}{i}\binom{n-1}{i-1}.$$

We evaluate the two inner sums as follows. Write the first one as

$$S_1 = \sum_{i=1}^{n}\binom{p}{i}\binom{n-1}{n-i} \qquad (p \geqq n).$$

Now, since

$$(1-t)^p = \binom{p}{0} + \binom{p}{1}t + \binom{p}{2}t^2 + \cdots + \binom{p}{p}t^p,$$

$$(1-t)^{n-1} = \binom{n-1}{0} + \binom{n-1}{1}t + \binom{n-1}{2}t^2 + \cdots + \binom{n-1}{n-1}t^{n-1},$$

the coefficient of t^n in the expansion of $(1-t)^{n+p-1}$ is

$$\binom{n-1}{n-1}\binom{p}{1} + \binom{n-1}{n-2}\binom{p}{2} + \cdots + \binom{n-1}{0}\binom{p}{n} = S_1.$$

On the other hand, the same coefficient is

$$\binom{n+p-1}{n} = S_1.$$

Now write the second sum as

$$S_2 = \sum_{i=1}^{p}\binom{p}{p-i}\binom{n-1}{i-1} \qquad (n > p).$$

The coefficient of t^{p-1} in the expansion of $(1-t)^{n+p-1}$ is, from the relations above,

$$\binom{p}{p-1}\binom{n-1}{0} + \binom{p}{p-2}\binom{n-1}{1} + \cdots + \binom{p}{0}\binom{n-1}{p-1} = S_2;$$

but the same coefficient is also

$$(53)\qquad \binom{n+p-1}{p-1} = \binom{n+p-1}{n} = S_2 = S_1.$$

Hence we have

$$(54)\quad \sum_{i=0}^{p} \binom{p}{i} |\Delta^{p-i}\mu_i| \leqq |\mu_\infty| + |\Delta^p \mu_0| + \sum_{n=1}^{\infty} |\Delta^p \mu_n| \binom{n+p-1}{p-1}$$

$$\leqq |\mu_\infty| + K \qquad (p = 1, 2, \cdots)$$

by (A). This shows that (A) implies (A′).

THEOREM 10b. *Condition* (A′) *implies condition* (A).

Let $p > 0$, arbitrary, be given. We then have

$$(1-t)^{k+1} t^{p-i} = (1-t)^{k+1} t^{p-i} (t + 1 - t)^{i-1}$$

$$= \sum_{j=0}^{i-1} \binom{i-1}{j} t^{p-j-1} (1-t)^{k+1+j} \qquad (0 < i \leqq p,\ k \geqq 0).$$

On each side of this equation we have a polynomial; since there are only a finite number of terms on each side, if we take the moments of both sides, using the given sequence $\{\mu_n\}$, we may conclude that the resulting moments are equal. Thus

$$(-1)^{k+1} \Delta^{k+1} \mu_{p-i} = \sum_{j=0}^{i-1} \binom{i-1}{j} (-1)^{k+j+1} \Delta^{k+j+1} \mu_{p-j-1},$$

and hence

$$|\Delta^{k+1} \mu_{p-i}| \leqq \sum_{j=0}^{i-1} \binom{i-1}{j} |\Delta^{k+j+1} \mu_{p-j-1}|.$$

Then, forming this relation for each i, $1 \leqq i \leqq p$, and adding, we obtain

$$\sum_{i=1}^{p} \binom{k+p-i}{k} |\Delta^{k+1}\mu_{p-i}| \leqq \sum_{i=1}^{p} \binom{k+p-i}{k} \sum_{j=0}^{i-1} \binom{i-1}{j} |\Delta^{k+j+1}\mu_{p-j-1}|$$

$$= \sum_{j=0}^{p-1} |\Delta^{k+j+1}\mu_{p-j-1}| \sum_{i=j+1}^{p} \binom{i-1}{j} \binom{k+p-i}{k}$$

$$= \sum_{j=0}^{p-1} \binom{p+k}{p-j-1} |\Delta^{k+j+1}\mu_{p-j-1}|.$$

For

$$\sum_{i=j+1}^{p} \binom{i-1}{j} \binom{k+p-i}{k}$$

is the coefficient of t^{p-j-1} in the Maclaurin series of $(1-t)^{-(j+1)}(1-t)^{-(k+1)} =$

$(1 - t)^{-(k+j+2)}$. Hence, writing in the last inequality i in place of $p - i$,

$$\sum_{i=0}^{p-1} \binom{k+i}{k} | \Delta^{k+1} \mu_i | \leqq \sum_{j=0}^{p-1} \binom{p+k}{p-j-1} | \Delta^{k+j+1} \mu_{p-j-1} |$$

$$= \sum_{m=0}^{p-1} \binom{p+k}{m} | \Delta^{p+k-m} \mu_m | \qquad (m = p - j - 1)$$

$$\leqq \sum_{m=0}^{p+k} \binom{p+k}{m} | \Delta^{p+k-m} \mu_m | \leqq K$$

by (A′). Since $p > 0$ was arbitrary, we have finally

$$\text{(A)} \qquad \sum_{i=0}^{\infty} \binom{i+k}{k} | \Delta^{k+1} \mu_i | \leqq K \qquad (k = 0, 1, 2, \cdots).$$

This completes the proof of the stated equivalence.

It is probably possible to show by similar methods that our conditions (B) and (E) are equivalent, in the same sense, to Hausdorff's conditions for the corresponding cases.

HARVARD UNIVERSITY.

GROUPS INVOLVING FIVE COMPLETE SETS OF NON-INVARIANT CONJUGATE OPERATORS

By D. T. Sigley

1. **Introduction.** The finite abstract groups involving no more than five complete sets of conjugate operators were determined by G. A. Miller.[1] W. Burnside[2] published the same results in his *Theory of Groups of Finite Order*. Some general theorems on the number of sets of conjugate operators in a group of finite order have been published by G. A. Miller.[3] In this paper we prove two theorems on complete sets of non-invariant conjugate operators in a group of finite order, and derive the abstract groups involving five complete sets of non-invariant conjugate operators.

2. **Non-invariant sets of conjugate operators in a finite group.** Let G represent a group of finite order g, and let H, of order h, represent the central of G. Assume that G is non-abelian, and hence G contains k, $k > 0$, complete sets of non-invariant conjugate operators. From the isomorphism between G and the quotient group G/H, we may state the

Lemma. *A necessary and sufficient condition that a group G contain more than k, $k > 0$, complete sets of non-invariant conjugate operators, if the central quotient group contains k complete sets of conjugate operators, and the identity, is that the order h, of H, exceed unity.*

We may state

Theorem 1. *A group G containing k, $k > 0$, complete sets of non-invariant conjugate operators, with a central of order greater than unity, has a central quotient group which involves not more than $k + 1$ complete sets of conjugate operators, at least two of which are composed of invariant operators, if this number is $k + 1$.*

The first part of the theorem follows from the fact that G and G/H are isomorphic, and hence G/H does not involve more complete sets of conjugates than G. Assume that G/H involves $k + 1$ complete sets of conjugate operators, of which k are composed of non-invariant operators. The order of G/H is divisible by at least two distinct primes p and q, since a group of order p^m, p a prime, contains invariant operators in addition to the identity. The operators in a co-set of G with respect to H are all conjugate. Therefore, all of the operators in the same co-set are of the same order. This is a contradiction, since there are operators of two different orders in one of the co-sets corresponding

Received February 18, 1935.

[1] Archiv der Math. und Phys., vol. 17 (1910), p. 199.

[2] W. Burnside, *Theory of Groups of Finite Order*, 2nd ed., 1911, Note A.

[3] Trans. Amer. Math. Soc., vol. 20 (1919), p. 262; Amer. Journal of Math., vol. 54 (1932), p. 110.

to an operator of order p or q in G/H. (Each operator of G is commutative with every operator of H.)

From the lemma and the theorem, it is seen that the number of complete sets of non-invariant conjugate operators in a group G is greater than the number of complete sets of non-invariant conjugate operators in the central quotient group of G.

Another result which has been useful in §3 is the

THEOREM 2. *A group G, having the non-cyclic group of order pq, p and q being the same or distinct primes with $p \geqq q$, as the central quotient group, contains $h(q^2 - q + p - 1)/q$ complete sets of conjugate non-invariant operators, where h is the order of the central of G.*

Establish an $(h - 1)$-isomorphism between G and the non-cyclic group of order pq, p and q primes with $p \geqq q$, so that the central of G corresponds to the identity of the group of order pq. The $(p - 1)h$ operators of G which correspond to operators of order p in G/H are distributed in $(p - 1)h/q$ conjugate sets of q operators each. The $p(q - 1)h$ operators of G which correspond to the operators of order q in G/H are distributed in $(q - 1)h$ conjugate sets of p operators each. The total number of non-invariant conjugate sets in G is $k = h(q^2 - q + p - 1)/q$.

COROLLARY. *If the central quotient group G/H is of order p^2, p a prime, the order of the central of G is divisible by p.*

For q equal to 2 or p in Theorem 2 the number of sets of conjugates in G is $h(p + 1)/2$, and $h(p^2 - 1)/p$, respectively.

3. **Groups involving five complete sets of non-invariant conjugate operators.** We make a division of cases as follows.

(A) Determination of the groups involving five complete sets of non-invariant conjugate operators, having a central whose order exceeds one, and

(B) Determination of the groups involving five complete sets of non-invariant conjugate operators, and having a central of order one.

We shall consider each case separately.

(A) From Theorem 1 the central quotient group in this case is one of the following groups.

(a) Non-cyclic group of order 4, 6, 10, or 14.

(b) Octic group.

(c) The tetrahedral group.

(d) The metacyclic group of order 20 or the semimetacyclic group of order 21.

(e) The symmetric group of order 24.

(f) The icosahedral group.

(g) The dihedral or dicyclic group of order 12.

As the consideration of these different cases is similar, we give only the determination for the cases (a) and (b).

(a) Substituting $p = 2, q = 2$, (the abelian G/H); $p = 3, q = 2$; $p = 5, q = 2$; and $p = 7, q = 2$ in the formula of Theorem 2, it is seen that no integral value of h gives five sets of conjugates.

(b) The operators of G corresponding to the operators of order four in G/H have only two conjugates. Hence they are in at least two sets. The operators of G corresponding to the characteristic operator of G/H are all conjugate. Therefore $h = 2$, and G is one of the groups of order 16 with the octic group as the central quotient group. The group G is, accordingly, the dihedral or dicyclic group of order 16, or $G_{16}(A^8 = B^2 = 1, BAB = A^3)$.

We consider next case

(B) The order of a group which involves five complete sets of non-invariant conjugate operators and a single invariant one satisfies an identity of the form[4]

$$1/g + 1/g_1 + 1/g_2 + 1/g_3 + 1/g_4 + 1/g_5 = 1\,,$$

where each g_i, $i = 1, 2, \cdots, 5$, is the order of a subgroup of G, and hence a divisor of g. An examination of the identities of this form reveals that the following are the only ones corresponding to groups involving five complete sets of non-invariant conjugate operators:

$$1/18 + 1/2 + 1/9 + 1/9 + 1/9 + 1/9 = 1\,, \tag{1}$$

$$1/168 + 1/3 + 1/4 + 1/8 + 1/7 + 1/7 = 1\,, \tag{2}$$

$$1/36 + 1/4 + 1/4 + 1/4 + 1/9 + 1/9 = 1\,. \tag{3}$$

In the study of these identities, we make the following observations. No g_i is equal to g, since there is a single invariant operator in G. If $p - 1$ of the g_i's are the same prime p, G contains an invariant subgroup of order prime to p. The number r of g_i's that are equal to a prime p is a divisor of $p - 1$, and $g/p \equiv (p - 1)/r$, mod p, for $r > 0$. If one of the g_i's is divisible by two distinct prime factors p and q, at least one other g_i is divisible by one of these prime numbers.

We may state the results in

Theorem 3. *The dihedral and generalized dihedral groups of order* 18, *the simple group of order* 168, *the dihedral and dicyclic groups of order* 16, *the non-twelve group of order* 24, *a group of order* 16 *with equations* $A^8 = B^2 = 1$, $BAB = A^3$, *and a group of order* 36 *with the equations* $A^3 = B^3 = C^4 = 1$, $AB = BA$, $C^{-1}AC = A^2B$, $C^{-1}BC = AB$ *are the only finite groups which involve five complete sets of non-invariant conjugate operators.*

University of Kansas City.

[4] Loc. cit., footnote 1.

ON SUBHARMONIC FUNCTIONS

By E. F. Beckenbach

1. **Introduction.** The following theorem, first given by Montel, was completed by Radó.[1]

A necessary and sufficient condition that the non-negative continuous function $p(u, v)$ be of class PL[2] is that for all real constants α, β the function

$$e^{\alpha u+\beta v}\, p(u, v)$$

be subharmonic.

The above theorem has been generalized by Kierst and Saks.[3]

It is the purpose of the present paper to present an immediately equivalent form (§2) of the Montel-Radó theorem, and to give two simple geometric consequences (§4 and §5). Without recourse to the Montel-Radó theorem, the latter of these consequent results has been given previously;[4] it is repeated briefly here because the present setting seems to be its proper one.

2. Lemma. *A necessary and sufficient condition that the non-negative continuous function $p(u, v)$, for (u, v) in some domain D, be of class PL is that for all analytic functions $f(u + iv)$, for (u, v) in D, the function*

$$p(u, v) \mid f(u + iv) \mid \tag{1}$$

be subharmonic.

Necessity. If $p(u, v)$ is of class PL, since the absolute value of an analytic function is of class PL, and since the product of two functions of class PL is a function of class PL, it follows that (1) is of class PL, and therefore, à fortiori, is subharmonic.

Sufficiency. If (1) is subharmonic for all $f(u + iv)$, it is subharmonic in

Received March 21, 1935.

[1] P. Montel, *Sur les fonctions convexes et les fonctions sousharmoniques*, Journal de Mathématiques, (9), vol. 7 (1928), pp. 29–60, especially, p. 40. T. Radó, *Remarque sur les fonctions subharmoniques*, Comptes Rendus, vol. 186 (1928), pp. 346–348. In proving the sufficiency, Montel assumed continuous partial derivatives of the first and second order; Radó removed this restriction.

[2] A function $p(u, v)$, defined in a domain D, is said to be of class PL in D provided $p(u, v)$ is continuous and $\geqq 0$ in D and $\log p(u, v)$ is subharmonic in the part of D where $p(u, v) > 0$. See E. F. Beckenbach and T. Radó, *Subharmonic functions and minimal surfaces*, Transactions of the American Mathematical Society, vol. 35 (1933), pp. 648–661 for the definition and elementary properties of these functions.

[3] S. Saks, *On subharmonic functions*, Acta Szeged, vol. 5 (1930–32), pp. 187–193.

[4] E. F. Beckenbach, *A characteristic property of surfaces of negative curvature*, Bull. Amer. Math. Soc., vol. 40 (1934), pp. 761–768.

particular for the function

(2) $$f(u + iv) = e^{(\alpha - i\beta)(u+iv)}, \qquad (u, v) \text{ in } D,$$

for arbitrary real α, β. We have then that

$$p(u, v) \mid e^{(\alpha - i\beta)(u+iv)} \mid \equiv p(u, v)\, e^{\alpha u + \beta v}$$

is subharmonic for all real constants α, β and hence, by the Montel-Radó theorem, that $p\,(u, v)$ is of class PL.

COROLLARY. *The lemma remains true if in* (1) *we replace* $f(u + iv)$ *by any power either of* $f(u + iv)$ *or of any derivative of* $f(u + iv)$, *with the restriction that for negative powers the function* (*either* $f(u + iv)$ *or the one of its derivatives in question*) *remains unequal to zero.*

The proof of the corollary is quite the same as that of the lemma, since in the proof of the sufficiency we used only functions (2) which remain unequal to zero.

3. Three continuous functions

(3) $$x = x(u, v), \qquad y = y(u, v), \qquad z = z(u, v), \qquad (u, v) \text{ in } D,$$

will be said to define a surface S.

If D is mapped conformally on an (r, s)-domain D^* by the analytic function

$$r + is = f(u + iv),$$

we have S in a representation

(4) $$x = x(u(r, s), v(r, s)) = X(r, s), \text{ etc.}, \qquad (r, s) \text{ in } D^*.$$

The representations (4), obtained from (3) by conformal maps of D on (r, s)-domains D^*, will be said to be *conformally equivalent* to the representation (3).

If the functions in (3) admit continuous first derivatives satisfying $E = G$, $F = 0$, where E, F, and G are the fundamental quantities of the first order for S, then u, v are said to be *isothermic* parameters, and the representation is conformal wherever $EG - F^2 \neq 0$. If u, v are isothermic parameters, so are r, s for every conformally equivalent representation.

If the representation (3) is isothermic, and if $x(u, v)$, $y(u, v)$, $z(u, v)$ are harmonic, these functions are said to be a *triple of conjugate harmonic functions*.[5] As is well known, a necessary and sufficient condition that (3) be an isothermic representation of a minimal surface is that $x(u, v)$, $y(u, v)$, $z(u, v)$ be a triple of conjugate harmonic functions.

4. THEOREM. *A necessary and sufficient condition that the continuous functions* $x(u, v)$, $y(u, v)$, $z(u, v)$ *in* (3) *be the coördinate functions of a minimal surface in isothermic representation is that for every conformally equivalent representation*

[5] See E. F. Beckenbach and T. Radó, loc. cit., p. 648.

(4) *with $r^2 + s^2 \neq 0$ the function*

$$\varphi(r, s) \equiv [(X - a)^2 + (Y - b)^2 + (Z - c)^2]^{\frac{1}{2}}/(r^2 + s^2)^{\frac{1}{2}}$$

be subharmonic for all real constants a, b, c.

Proof. We have

$$\varphi(r, s) = [(x - a)^2 + (y - b)^2 + (z - c)^2]^{\frac{1}{2}}/|f(u + iv)| = \psi(u, v),$$

whence by the corollary of §2 the function

$$[(x - a)^2 + (y - b)^2 + (z - c)^2]^{\frac{1}{2}} \tag{5}$$

is of class PL. But the condition that (5) be of class PL for all real a, b, c is necessary and sufficient in order that $x(u, v)$, $y(u, v)$, $z(u, v)$ be a triple of conjugate harmonic functions, that is, that $x(u, v)$, $y(u, v)$, $z(u, v)$ be the coördinate functions of a minimal surface in isothermic representation.[6]

The theorem remains true, by the lemma of §2, if in place of $\varphi(r, s)$, representing the quotient of distances with $r^2 + s^2 \neq 0$, we substitute the function

$$[(X - a)^2 + (Y - b)^2 + (Z - c)^2]^{\frac{1}{2}}(r^2 + s^2)^{\frac{1}{2}},$$

representing the product of distances.

5. Theorem. *If the functions $x(u, v)$, $y(u, v)$, $z(u, v)$ in (3) have continuous derivatives of the third order, and if u, v are isothermic parameters, a necessary and sufficient condition that the Gaussian curvature K of the surface S be $\leqq 0$ wherever K is defined[7] on S is that, for every conformally equivalent representation, the area deformation ratio*

$$\mu(r, s) \equiv X_r^2 + Y_r^2 + Z_r^2$$

be subharmonic.

Proof. We have

$$\mu(r, s) = (x_u^2 + y_u^2 + z_u^2)/|f'(u + iv)|^2 = \lambda(u, v)/|f'(u + iv)|^2,$$

where $\lambda(u, v) \geqq 0$. By §2, then, a necessary and sufficient condition that $\lambda(u, v)$ be of class PL is that $\mu(r, s)$ be subharmonic for all conformally equivalent representations. Since $\lambda(u, v)$ has continuous second derivatives, the statement that $\lambda(u, v)$ is of class PL is equivalent to the statement that

$$\Delta \log \lambda \equiv \frac{\partial^2 \log \lambda}{\partial u^2} + \frac{\partial^2 \log \lambda}{\partial v^2} \geqq 0$$

in the part of D where $\lambda > 0$; further, K is defined at points where $\lambda > 0$ and is then given by

$$K = -\frac{1}{2\lambda} \Delta \log \lambda;$$

[6] See E. F. Beckenbach and T. Radó, loc. cit., p. 654.

[7] I.e., wherever $EG - F^2 \neq 0$.

consequently the two conditions, (i) that λ be of class PL, and (ii) that $K \leqq 0$ in the part of D where K is defined, are equivalent. The theorem follows from these considerations.

In the statement of the present theorem, we could as well have used the length deformation ratio $[\mu(r, s)]^{\frac{1}{2}}$ in place of the area deformation ratio $\mu(r, s)$. For since the square of a subharmonic function is subharmonic, it follows that if $[\mu(r, s)]^{\frac{1}{2}}$ is subharmonic, so is $\mu(r, s)$. Conversely, if $\mu(r, s)$ is subharmonic for all conformally equivalent representations, $\lambda(u, v)$ is of class PL. Since the class PL is invariant under conformal mappings, $\mu(r, s)$ is of class PL. Finally, since any positive power of a function of class PL is subharmonic, $[\mu(r, s)]^{\frac{1}{2}}$ is subharmonic.

THE RICE INSTITUTE.

ON THE NUMBER THEORY OF CERTAIN NON-MAXIMAL DOMAINS OF THE TOTAL MATRIC ALGEBRA OF ORDER 4

By Edward J. Finan

1. **Introduction.** This paper is devoted to the investigation of the number theory of certain non-maximal domains of integrity of the total matric algebra of order 4.

We shall call a *domain of integrity* (or merely a *domain*) of the above algebra any subset which (1) is of order 4, (2) contains the identity matrix and (3) is closed under addition and multiplication, the constants of multiplication being rational integers.

A[1] canonical basis has been derived for such domains under certain transformations. We shall make a study of a subset of these domains, obtaining some interesting properties. If we take the basis mentioned above under case I and set $m = l = 0$, $a = 1$ and let k be a prime, we get a basis which is evidently equivalent to

$$(1) \qquad \left\|\begin{matrix}1 & 0\\ 0 & 0\end{matrix}\right\|, \quad \left\|\begin{matrix}0 & 1\\ 0 & 0\end{matrix}\right\|, \quad \left\|\begin{matrix}0 & 0\\ 0 & 1\end{matrix}\right\|, \quad \left\|\begin{matrix}0 & 0\\ k & 0\end{matrix}\right\|.$$

We shall refer to the above matrices as E_1, E_2, E_3, and E_4 in the order given. If $k = 1$, we get the unique maximal domain of the algebra. In this paper $|k| > 1$. We shall refer to (1) as the domain D.

In paragraph 2 we obtain a set of canonical forms for the numbers of the domain under consideration. From this it follows that a necessary and sufficient condition that a number be indecomposable in the domain is that its determinant be a rational prime. Hence we have an example of a simple non-commutative domain of class number greater than one for which the indecomposable numbers are known.

Paragraph 3 is devoted to the determination of the class number of the domain. I believe this is the first determination of the class number of a non-commutative domain with class number greater than 1—that of quaternions being 1. This approach to the theory of ideals through matrices with rational integral elements is the same as that used by C. C. MacDuffee.[2]

2. **Canonical forms for numbers of the domain.** Any number in the domain D may be written in the form $N = \Sigma\, n_i E_i$, where the n_i are rational integers. Evidently such a number is a rational integral square matrix of

Received April 8, 1935.

[1] E. J. Finan, American Journal of Mathematics, vol. 54 (1931), pp. 920–928.

[2] C. C. MacDuffee, Transactions of the Amer. Math. Soc., vol. 31 (1929), pp. 71–90.

order 2. We shall call N a unit if and only if $|N| = \pm 1$. Two numbers N_1 and N_2 in D are said to be associated if and only if there exist in D units A and B such that $AN_1B = N_2$. We might mention here that the relations $n_1 \equiv 0$, $n_3 \equiv 0 \bmod k$ and every scalar factor are invariants in all associated numbers of D. Since the inverse of a unit in D is a unit in D, the above relationship is reciprocal. It is easily shown that the centrum[3] of D is the set of all two-rowed scalar matrices. Hence in finding a canonical form for associated numbers we shall neglect the scalar factors. We shall now prove that every number of D without a scalar factor other than a unit is associated with one and only one (with the exceptions stated in the theorem) of the numbers

$$\begin{Vmatrix} 1 & 0 \\ 0 & n_3' \end{Vmatrix}, \quad \begin{Vmatrix} n_1' & 0 \\ 0 & 1 \end{Vmatrix}, \quad \begin{Vmatrix} 0 & n_2' \\ k & 0 \end{Vmatrix}, \quad \begin{Vmatrix} 0 & 1 \\ kn_4' & 0 \end{Vmatrix},$$

which we shall refer to as α, β, γ, and δ, respectively. The n''s in the above canonical forms are rational integers and k is the rational prime in (1).

THEOREM 1. *Let $N = \Sigma n_i E_i$ be any number of D, without a scalar factor $s = sE_1 + sE_3$, whose basis is (1). Then N is associated with α, β, γ, or δ, according as $n_1 \not\equiv 0$, $n_3 \not\equiv 0$, $(n_1, n_2, n_3) \equiv (0, 0, 0)$, $(n_1, n_3, n_4) \equiv (0, 0, 0) \bmod k$, respectively. If $(n_1, n_3) \equiv (0, 0)$ and $n_2 n_4 \not\equiv 0 \bmod k$, N is associated with both δ and γ. If $n_1 n_3 \not\equiv 0 \bmod k$, N is associated with both α and β. Finally, if $n_1' n_3' \not\equiv 0 \bmod k$, α and β are associated. If $n_2' \not\equiv 0 \bmod k$, γ and δ are associated. In all other cases no two of the four numbers in canonical form are associated.*

We shall first prove that if $n_1 \not\equiv 0 \bmod k$, N is associated with α. We make two cases of the proof.

Case I. n_2 contains as many factors k as n_3. Form the product $AN = N' = \Sigma n_i' E_i$ where $A = \Sigma a_i E_i$. Then $n_2' = a_1 n_2 + a_2 n_3$. Since n_2 contains as many factors k as n_3, there exist values for a_1 and a_2 such that n_2' is zero and such that a_1 and ka_2 will be relatively prime. Then it is possible to select a_3 and a_4 so that $|A| = a_1 a_3 - ka_2 a_4 = 1$. Hence N is associated with a number N' for which $n_2' = 0$. Now form the product $CN'B = N'' = \Sigma n_i'' E_i$ where $C = \Sigma c_i E_i$ and $B = \Sigma b_i E_i$. The element in the upper left hand corner of N'' is

$$n_1'' = b_1 c_1 n_1' + b_1 c_2 k n_4' + c_2 b_4 k n_3'. \tag{2}$$

Since N has no scalar factor, n_1', n_3' and n_4' are relatively prime. Also, since A is a unit, $n_1' \not\equiv 0 \bmod k$. Hence n_1', kn_4', and kn_3' are relatively prime. Then[4] there exist integers b_1, c_1, c_2, and b_4 such that $n_1'' = 1$. However, since N' and N'' are to be associated, we must show that the conditions $|B| = |C| = \pm 1$ can be satisfied at the same time. Since $n_1'' = 1$, it is evident from (2) that c_1 cannot have a factor k and that c_1 and c_2 must be relatively prime. Hence c_1 and kc_2 are relatively prime and we can find integers c_3 and c_4 which will satisfy $|C| = 1$.

[3] The *centrum* is the set of elements of D commutative with every element of D. See *Moderne Algebra* by van der Waerden, vol. II, p. 163.

[4] E. J. Finan, Bulletin of the Amer. Math. Soc., vol. 39 (1933), p. 944.

In the same way we can show the existence of integral values of b_3 and b_2 that will satisfy $|B| = 1$.

Case II. n_3 contains more factors k than n_2. The procedure here is similar to that in Case I. We start out by making $n_3' = 0$ and then proceed as above until we get $n_1'' = 1$. Dropping primes, we have in both Case I and Case II

$$N = \begin{Vmatrix} 1 & n_2 \\ kn_4 & n_3 \end{Vmatrix}.$$

If we multiply on the left and right respectively by the units

$$\begin{Vmatrix} 1 & 0 \\ -kn_4 & 1 \end{Vmatrix}, \qquad \begin{Vmatrix} 1 & -n_2 \\ 0 & 1 \end{Vmatrix},$$

we get the α mentioned in the theorem.

The derivation of the remaining canonical forms given in the theorem is similar to the above. We shall not include it here.

We shall now prove the second part of the theorem. If the n_1' in β is not divisible by k, then β is associated with α according to the proof given above. To show that γ and δ are associated when $n_2' \not\equiv 0 \bmod k$, form the product $N'' = A\gamma B$. The element in the upper right hand corner is $n_2'' = b_2a_2k + a_1b_3n_2'$, which can be made unity by proper choice of a_1, a_2, b_2, and b_3. Values for a_3, a_4, b_1, and b_4 may then be chosen so that $|A| = |B| = 1$. If $n_2'' = 1$ and $(n_1'', n_3'') \equiv (0, 0)$, N'' and δ are associated.

To prove the last statement in the theorem, we observe that if $n_2' \equiv 0 \bmod k$, then $n_2'' = b_2a_2k + a_1b_3n_2'$ can not be unity. Hence γ and δ are not associated. In view of the invariant relation given above, β is not associated with any of the remaining canonical forms providing $n_1' \equiv 0 \bmod k$. A similar remark holds for α. This completes the proof of the theorem.

THEOREM 2. *A necessary and sufficient condition that N be a prime is that $|N|$ be a rational prime.*

The sufficiency of the condition is evident. Suppose $|N|$ is not a prime. If N is the scalar s, it is the product of $S_1 = E_1 + sE_3$ and $S_2 = sE_1 + E_3$, neither of which is a unit. If N has no scalar factor, it is associated with α, β, γ, or δ according to Theorem 1. For example, suppose N is associated with γ, and

$$ANB = \begin{Vmatrix} 0 & n_2 \\ k & 0 \end{Vmatrix} = \begin{Vmatrix} 0 & 1 \\ k & 0 \end{Vmatrix} \cdot \begin{Vmatrix} 1 & 0 \\ 0 & n_2 \end{Vmatrix},$$

where A and B are units. Multiplying the above matric equation on the left and right by A^{-1} and B^{-1}, respectively, we have N expressed as the product of two numbers, neither of which is a unit. In a similar way, it can be shown that N is factorable if it is associated with α, β, or δ.

3. **Ideals in the domain.** We shall define a left ideal[5] of D to be a set of numbers of D which is closed under addition and subtraction and under multi-

[5] C. C. MacDuffee, Transactions of Amer. Math. Soc., vol. 31, pp. 71–90.

plication on the left by numbers of D. It has been shown that such an ideal has a basis ω_1, ω_2, ω_3, and ω_4, where $\omega_i = \Sigma g_{ij}E_j$. A[6] necessary and sufficient condition that the ω's constitute a basis is that there exist integral matrices D_1, D_2, D_3, and D_4 such that

$$GC_p = D_pG \qquad (p = 1, 2, 3, 4), \tag{3}$$

where G is the matrix (g_{rs}) above defining the ω's, and $C_p = c_{prs}$, the c's being the constants of multiplication of the domain. We shall refer to G as an ideal matrix, or briefly as an ideal.

A[7] necessary and sufficient condition that the ideals G_1 and G_2 be equivalent is that the corresponding sets of matrices D_{1p}, D_{2p}, D_{3p}, and D_{4p} satisfying the equations

$$G_1C_p = D_{1p}G_1, \qquad G_2C_p = D_{2p}G_2 \qquad (p = 1, 2, 3, 4),$$

respectively, be similar—i.e., that $D_{1p} = AD_{2p}A^{-1}$ for $p = 1, 2, 3, 4$, where A is an integral unit matrix. The left class number of D is the number of non-equivalent non-singular ideals in D. It is known to be finite.[8] We shall show that this class number is 3. Because of the length of the calculations we shall give an outline only. If G is an ideal and A is a unit integral matrix, then G and AG stand for the same ideal. Hence G may be taken as an integral non-singular matrix of order 4 in Hermite's canonical form.[9] G is taken as non-singular[10] because G is a non-singular ideal if and only if G is a non-singular matrix. The matrices

$$\begin{Vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{Vmatrix}, \quad \begin{Vmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ k & 0 & 0 & 0 \end{Vmatrix}, \quad \begin{Vmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{Vmatrix}, \quad \begin{Vmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & k & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{Vmatrix}$$

are the C_1, C_2, C_3 and C_4 respectively of (3). Now form the matrices GC_pG^{-1} $(p = 1, 2, 3, 4)$. The condition that these 4 matrices be integral together with the condition that G be in Hermite's canonical form show that G must be in the form

$$G' = \begin{Vmatrix} \alpha g_{44} & 0 & 0 & 0 \\ \beta g_{44} & g_{22} & 0 & 0 \\ 0 & 0 & \gamma g_{22} & 0 \\ 0 & 0 & \delta g_{22} & g_{44} \end{Vmatrix},$$

[6] C. C. MacDuffee, loc. cit., p. 76.
[7] C. C. MacDuffee, loc. cit., p. 79.
[8] C. G. Latimer, Bulletin of the Amer. Math. Soc., vol. 40 (1934), p. 433.
[9] Journal für Mathematik, vol. 41 (1851), p. 193.
[10] C. C. MacDuffee, loc. cit., p. 74.

where

(4) $\beta\delta \equiv k \bmod \alpha$ and $\bmod \gamma$, $\quad \alpha\delta \equiv 0 \bmod \gamma$, $\quad \beta\gamma \equiv 0 \bmod \alpha$.

If one forms the D_1, D_2, D_3, and D_4 of (3) using $G = G'$, one finds them identical with those obtained by using G equal to

$$G'' = \begin{Vmatrix} \alpha & 0 & 0 & 0 \\ \beta & 1 & 0 & 0 \\ 0 & 0 & \gamma & 0 \\ 0 & 0 & \delta & 1 \end{Vmatrix},$$

where α, β, γ, and δ satisfy (4). Hence every ideal is equivalent to one in the form G'', where (4) still holds.

We shall now show that G'' is equivalent to a diagonal ideal matrix. Let $\alpha = f\alpha_1, \gamma = f\gamma_1$, where α_1 and γ_1 are relatively prime. Then because of the last two congruences in (4), β contains the factor α_1 and δ contains γ_1. From the first two congruences k contains α_1 and γ_1, but since they are relatively prime, k contains $\alpha_1\gamma_1$. The only possibilities are (a) $\alpha_1 = \pm k$ and $\gamma_1 = \pm 1$; (b) $\alpha_1 = \pm 1$ and $\gamma_1 = \pm k$; (c) $\alpha_1 = \pm 1$ and $\gamma_1 = \pm 1$. Evidently the discussion is general if we consider only the positive signs in the above equations. In all three cases G'' is equivalent to

$$G''' = \begin{Vmatrix} \alpha_1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \gamma_1 & 0 \\ 0 & 0 & 0 & 1 \end{Vmatrix}.$$

We shall give the proof for the first case only—i.e., we shall show that the ideals represented by

$$G'' = \begin{Vmatrix} fk & 0 & 0 & 0 \\ \beta_1 k & 1 & 0 & 0 \\ 0 & 0 & f & 0 \\ 0 & 0 & \delta & 1 \end{Vmatrix} \quad \text{and} \quad G''' = \begin{Vmatrix} k & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{Vmatrix}$$

are equivalent. To do this, we shall find a matrix $A = (a_{rs})$ such that $|A| = \pm 1$ and $AG''C_pG''^{-1} = G'''C_pG'''^{-1}A$ for $p = 1, 2, 3$, and 4. Since β contains $\alpha_1 = k$, we can write $\beta = \beta_1 k$. Also in this case (4) is equivalent to the single condition $1 - \beta_1\delta \equiv 0 \bmod f$. Hence let $1 - \beta_1\delta = ef$. When $p = 1$ or 3, the condition is satisfied by the matrix which is determined from the remaining two matrix equations. When $p = 2$ and 4, we get 8 equations—4 for each value of p:

$$
\begin{aligned}
-\beta_1 a_{33} + e a_{34} &= a_{21}, & -\delta a_{11} + e a_{12} &= a_{43},\\
f a_{33} + \delta a_{34} &= a_{22}, & f a_{11} + \beta_1 a_{12} &= a_{44},\\
-\beta_1 a_{43} + e a_{44} &= a_{11}, & -\delta a_{21} + e a_{22} &= a_{33},\\
f a_{43} + \delta a_{44} &= a_{12}, & f a_{21} + \beta_1 a_{22} &= a_{34}.
\end{aligned}
$$

If we solve the last 4 equations in pairs for a_{11}, a_{12}, a_{21} and a_{22}, we see that they are equivalent to the first 4. Hence assign values to a_{11}, a_{12}, a_{21} and a_{22} so that the two-rowed determinant in the upper left-hand corner of A is 1. Since the determinant of the unknowns in the first two equations is a unit, we get integral values for a_{33} and a_{34}. The same is true for a_{43} and a_{44}. Hence we have shown the existence of an integral unit matrix A, whose elements satisfy the 8 equations above. Thus G'' and G''' are equivalent. The simplest matrix satisfying these equations is the first one of the three given below.

$$
\begin{Vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & e & \beta_1 \\ 0 & 0 & -\delta & f \end{Vmatrix}, \quad
\begin{Vmatrix} g & \delta_1 & 0 & 0 \\ -\beta & f & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{Vmatrix}, \quad
\begin{Vmatrix} a_{11} & a_{12} & 0 & 0 \\ -\beta & \alpha & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & ha_{12} - \delta a_{11} & 1 \end{Vmatrix}.
$$

The second matrix is the A matrix for case (b) where conditions (4) become $1 - \delta_1\beta = gf$ and $\delta = \delta_1 k$. The third matrix is the A matrix for case (c) with α prime to both β and δ. The conditions (4) then become $k - \beta\delta = h\alpha$. In this matrix a_{11} and a_{12} are any integers such that $\alpha a_{11} + \beta a_{12} = 1$. There are two more possibilities under case (c), but we shall not include the A matrices for them.

Hence all ideals are equivalent to

$$
\begin{Vmatrix} k & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{Vmatrix}, \quad
\begin{Vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & k & 0 \\ 0 & 0 & 0 & 1 \end{Vmatrix}, \quad
\begin{Vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{Vmatrix},
$$

which we shall refer to as G_1, G_2 and G_3 respectively. To show that these three are non-equivalent, it is sufficient to form $G_iC_2G_i^{-1}$ $(i = 1, 2, 3)$. These matrices are respectively

$$
\begin{Vmatrix} 0 & & 0 \\ 0 & 1 & \\ 1 & 0 & 0 \end{Vmatrix}, \quad
\begin{Vmatrix} 0 & & 0 \\ 0 & k & \\ k & 0 & 0 \end{Vmatrix}, \quad
\begin{Vmatrix} 0 & & 0 \\ 0 & 1 & \\ k & 0 & 0 \end{Vmatrix}.
$$

If $AG_1C_2G_1^{-1} = G_2C_2G_2^{-1}A$, we have

$$\begin{Vmatrix} a_{14} & a_{13} & 0 & 0 \\ a_{24} & a_{23} & 0 & 0 \\ a_{34} & a_{33} & 0 & 0 \\ a_{44} & a_{43} & 0 & 0 \end{Vmatrix} = \begin{Vmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ ka_{21} & ka_{22} & ka_{23} & ka_{24} \\ ka_{11} & ka_{12} & ka_{13} & ka_{14} \end{Vmatrix}.$$

This implies that $(a_{14}, a_{24}, a_{34}, a_{44}) \equiv (0, 0, 0, 0) \bmod k$. This is impossible if $| A | = \pm 1$. Hence G_1 and G_2 are non-equivalent. In the same way, it may be shown that G_1 and G_3 are non-equivalent, and that G_2 and G_3 are non-equivalent. Hence the left class number of D is three.

It is well known[11] that a necessary and sufficient condition that every pair of numbers of D have a greatest common right divisor expressible linearly in terms of the numbers is that the left class number be 1. Hence there are pairs of numbers in D having no g.c.r.d. Consider the 4 numbers

$$\begin{Vmatrix} k & 0 \\ 0 & k \end{Vmatrix}, \quad \begin{Vmatrix} 0 & k \\ k & 0 \end{Vmatrix}, \quad \begin{Vmatrix} 1 & 0 \\ 0 & k \end{Vmatrix}, \quad \begin{Vmatrix} 0 & 1 \\ k & 0 \end{Vmatrix}.$$

The last two are right divisors of the first two. However, they are primes by Theorem 2 and non-associates by Theorem 1. Hence the first two numbers do not possess a greatest common right divisor.

THE CATHOLIC UNIVERSITY OF AMERICA.

[11] C. C. MacDuffee, loc. cit., p. 89.

CONCERNING THE EQUILIBRIUM POINT OF GREEN'S FUNCTION FOR AN ANNULUS

By Alfred J. Maria

1. Introduction. This paper is concerned with an investigation of the equilibrium point of Green's function for an annulus.

In the z-plane let $|z| = R_1$ and $|z| = R$, $R_1 < R$, be the equations of two circles; let Σ be the set of points z for which $R_1 \leqq |z| \leqq R$. It is known that the Green's function for Σ, $G(z, z_0)$, with pole at z_0, interior to Σ, has one and only one equilibrium point z, $R_1 < |z| < R$, i.e., $z = re^{i\theta}$ satisfies the equations $\partial G/\partial r = 0$, $\partial G/\partial \theta = 0$; further, $\theta = \theta_0 + \pi$.[1]

Let the pole z_0 lie on a fixed line σ through 0; then the equilibrium point z will lie on σ. We shall then have r as a single-valued function of r_0, $r = g(r_0)$, say. The following is a study of the function $g(r_0)$.

It will be shown that $g'(r_0) > 0$, $R_1 < r_0 < R$, and that $\lim_{r_0 \to R} g' = \lim_{r_0 \to R_1} g' = 0$. Another result is that if $R_1 < r_0 < R$, values R_2, R_3 exist such that $R_1 < R_2 < r < R_3 < R$. It is possible from the theory developed to compute for given R_1 and R the maximum and minimum values of R_2 and R_3, respectively.

2. Some results in the theory of elliptic functions.[2] The function $\zeta(u)$ is defined as

$$\zeta(u) = \frac{1}{u} + \sum_{\omega} \left(\frac{1}{u - 2\omega} + \frac{1}{2\omega} + \frac{u}{4\omega^2} \right), \qquad \omega = m_1\omega_1 + m_3\omega_3 \neq 0.$$

The function $\sigma(u)$ is defined as

$$\sigma(u) = \exp\left(\log u + \sum_{\omega} \left\{ \log\left(1 - \frac{u}{2\omega}\right) + \frac{u}{2\omega} + \frac{u^2}{8\omega^2} \right\} \right).$$

The function $\wp(u) = -\zeta'(u)$ is doubly periodic with periods $2\omega_1$ and $2\omega_3$. $\zeta(u)$ and $\sigma(u)$ are odd functions, and $\wp(u)$ is an even function. When ω_1 is positive and ω_3 is pure imaginary, $\zeta(u)$, $\sigma(u)$ and $\wp(u)$ are real for real u. We define $\eta_1 = \zeta(\omega_1)$ which is real.

3. **The Green's function for Σ.** It is clear that we may restrict the discussion, without loss of generality, to the case where $R_1 = 1$ and the pole is at e^{α}, α

Received April 26, 1935; presented to the American Mathematical Society June 15, 1933.

[1] For the sake of completeness, proofs of these statements will be given.

[2] Tannery et Molk, *Eléments de la théorie des fonctions elliptiques*, Paris, 1893, vol. 2, p. 235.

positive. Under these restrictions it can be easily shown with the results of section 2 that $G(z, z_0) \equiv \Re \log V(\log z)$, where

$$V(u) = \frac{\sigma(u + \alpha)}{\sigma(u - \alpha)} \exp(-2\lambda\alpha u),$$

$\omega_1 = \log R$, $\lambda = \eta_1/\omega_1$, $\omega_3 = \pi i$ and $u = \log z$, is the Green's function for Σ.[3]

It is clear that the equilibrium points for $G(z, z_0)$ are the solutions of

$$\frac{d \log V(u)}{du} = -2\lambda\alpha - \zeta(u - \alpha) + \zeta(u + \alpha) = 0. \tag{1}$$

It can be shown that $d \log V(u)/du = 0$ for $u = \omega_3 + t$, $0 < t < \omega_1$. Now $\phi_\alpha(u) = \zeta(u + \alpha) - \zeta(u - \alpha)$ is an even doubly-periodic function with primitive periods $2\omega_1$, $2\omega_3$, since[4]

$$\phi_\alpha(u) = 2\zeta(\alpha) - \frac{\wp'(\alpha)}{\wp(u) - \wp(\alpha)}.$$

Hence there are just two values of u in the period parallelogram with vertices at $\omega_1 + \omega_3$, $\omega_1 - \omega_3$, $-\omega_1 - \omega_3$, $-\omega_1 + \omega_3$, where $\phi_\alpha(u) = 2\lambda\alpha$; $u = -\omega_3 - t$, $u = -\omega_3 + t$. It follows then that $G(z, z_0)$ has exactly one equilibrium point in Σ and that it lies on the same diameter as the pole.

4. **The derivative** $g'(r_0)$. To show that $g' \neq 0$, it is sufficient to show that $du/d\alpha \neq 0$. From (1) we have

$$0 = 2\lambda + \wp(u - \alpha) + \wp(u + \alpha) + \frac{du}{d\alpha}[\wp(u + \alpha) - \wp(u - \alpha)], \tag{2}$$

where α is such that $0 < \alpha < \omega_1$ and $u = \omega_3 + t$, $0 < t < \omega_1$ is the solution of

$$2\lambda\alpha + \zeta(u - \alpha) - \zeta(u + \alpha) = 0. \tag{3}$$

In the first place $\wp(u + \alpha) - \wp(u - \alpha) \neq 0$ and is finite for $u = \omega_3 + t$ and $0 < \alpha < \omega_1$, since

$$\wp(u + \alpha) - \wp(u - \alpha) = -\frac{\wp'(u)\wp'(\alpha)}{[\wp(u) - \wp(\alpha)]^2}.$$

Using the identities[5]

$$\zeta(s + t) - \zeta(s) - \zeta(t) = \frac{1}{2}\,\frac{\wp'(s) - \wp'(t)}{\wp(s) - \wp(t)},$$

$$\wp(s + t) + \wp(s) + \wp(t) = \frac{1}{4}\left(\frac{\wp'(s) - \wp'(t)}{\wp(s) - \wp(t)}\right)^2$$

[3] Goursat, *Cours d'analyse mathématique*, Paris, 1915, vol. 3, p. 241.

[4] Tannery et Molk, loc. cit., vol. 4, p. 96.

[5] Tannery et Molk, loc. cit., vol. 4, p. 96.

and letting $s + t = u - \alpha$, $s = u + \alpha$, we must have

$$[\zeta(2\alpha) - 2\lambda\alpha]^2 = -2\lambda + \wp(2\alpha)$$

for at least one value of α such that $0 < \alpha < \omega_1$ if

$$2\lambda + \wp(u - \alpha) + \wp(u + \alpha) = 0,$$

where u and α satisfy (3). It will be shown that this is impossible.

Let $2\alpha = y$. We have

$$[\zeta(y) - \lambda y]^2 = -2\lambda + \wp(y). \tag{4}$$

It will be sufficient to show that this does not hold for y real and such that $0 < y \leqq \omega_1$. In fact, if

$$2\lambda\alpha + \zeta(\omega_3 + t - \alpha) - \zeta(\omega_3 + t + \alpha) = 0$$

and

$$2\lambda + \wp(\omega_3 + t - \alpha) + \wp(\omega_3 + t + \alpha) \neq 0,$$

then

$$2\lambda(\omega_1 - \alpha) + \zeta(\omega_1 + \omega_3 - t - (\omega_1 - \alpha)) - \zeta(\omega_1 + \omega_3 - t + (\omega_1 - \alpha)) = 0$$

and

$$2\lambda + \wp(\omega_1 + \omega_3 - t - (\omega_1 - \alpha)) + \wp(\omega_1 + \omega_3 - t + (\omega_1 - \alpha)) \neq 0.$$

This is easily demonstrated, using merely simple properties of $\zeta(u)$ and $\wp(u)$.

More results in the theory of elliptic functions.[6] We take $\omega_1 = \log R$, $\omega_3 = \pi i$. Then $\wp(\omega_1) = e_1$, $\wp(\omega_2) = \wp(-\omega_1 - \omega_3) = e_2$ and $\wp(\omega_3) = e_3$ are real, and $e_1 > e_2 > e_3$; e_1, e_2 and e_3 are the zeros of $4x^3 - g_2x - g_3$. The relations $g_2 = 4(e_1^2 - e_2e_3) = 2(e_1^2 + e_2^2 + e_3^2)$, $g_3 = 4e_1e_2e_3$, $0 = e_1 + e_2 + e_3$ are seen to hold.

The following formulas hold:

$$2\eta_1\omega_1 = -2e_1\omega_1^2 + \pi^2\left[\frac{1}{2} + \sum_{n=1}^{\infty} \frac{4q^{2n}}{(1 + q^{2n})^2}\right], \tag{5}$$

$$2\eta_1\omega_1 = -2e_2\omega_1^2 + \pi^2 \sum_{n=1}^{\infty} \frac{4q^{2n-1}}{(1 + q^{2n-1})^2}, \tag{6}$$

$$2\eta_1\omega_1 = -2e_3\omega_1^2 - \pi^2 \sum_{n=1}^{\infty} \frac{4q^{2n-1}}{(1 - q^{2n-1})^2}, \tag{7}$$

where $q = \exp(i\omega_3/\omega_1)$ and $\eta_1 = \zeta(\omega_1)$.

[6] Tannery et Molk, loc. cit., vol. 2, pp. 236 and 250.

From (6) and (7) it is seen that

$$e_2 + \lambda < -e_3 - \lambda.$$

Hence $2\lambda < -e_2 - e_3 = e_1$. Therefore $\wp(y) - 2\lambda > 0$ for y in $I(0 < y \leqq \omega_1)$. Furthermore, $\zeta(y) - \lambda y > 0$ for $0 < y < \omega_1$, because of (5) and the relations

$$\frac{d}{dy}(\zeta(y) - \lambda y) = -\wp(y) - \lambda < 0, \qquad \zeta(\omega_1) - \eta_1 = 0, \qquad \zeta(y) - \lambda y > 0,$$

which hold for real positive sufficiently small y.

If (4) holds for some y in I, then

$$\zeta(y) - \lambda y = \sqrt{\wp(y) - 2\lambda}.$$

From the series expansions for $\zeta(y)$ and $\wp(y)$ it follows that

$$\lim_{y \to 0} \left[\zeta(y) - \lambda y - \sqrt{\wp(y) - 2\lambda}\right] = 0.$$

Therefore if (4) holds for some y in I, there is a y in I for which

$$\frac{\wp'(y)}{2\sqrt{\wp(y) - 2\lambda}} = -\wp(y) - \lambda. \tag{8}$$

This is elementary.

It follows that we shall have

$$\wp(y) = \frac{8\lambda^3 - g_3}{g_2 - 12\lambda^2},$$

when we square both sides of (8) and use the identity $[\wp'(y)]^2 = 4[\wp(y)]^3 - g_2\wp(y) - g_3$. Since $\wp(y)$, y in I, is not less than e_1, the inequality

$$\frac{8\lambda^3 - g_3}{g_2 - 12\lambda^2} \geqq e_1 \tag{9}$$

must hold.

Let us show that $g_2 - 12\lambda^2 > 0$. Suppose $\lambda > 0$; then $e_1 > 2\lambda > 0$. If $e_2 \geqq 0$,

$$g_2 = 4(e_1^2 - e_2e_3) \geqq 4e_1^2 > 16\lambda^2 > 12\lambda^2;$$

if $e_2 < 0$,

$$g_2 = 4(e_1^2 - e_2e_3) = 4(e_1^2 - a(1 - a)e_1^2) \geqq 3e_1^2 > 12\lambda^2.$$

If $\lambda \leqq 0$, it follows from (5) and (6) that

$$e_1 > |\lambda|, \qquad e_2 > |\lambda|.$$

Hence

$$-e_3 = |e_3| = e_1 + e_2 > 2|\lambda|.$$

Therefore

$$2(e_1^2 - \lambda^2) + 2(e_2^2 - \lambda^2) + 2(e_3^2 - 4\lambda^2) = 2(e_1^2 + e_2^2 + e_3^2) - 12\lambda^2$$
$$= g_2 - 12\lambda^2 > 0.$$

Multiplying both sides of (9) by $g_2 - 12\,\lambda^2$ we obtain

$$8\lambda^3 - g_3 \geqq e_1 g_2 - 12\lambda^2 e_1. \tag{10}$$

Now $g_2 = 4(e_1^2 - e_2 e_3)$ and $g_3 = 4e_1e_2e_3$. Substituting these values into (10), we get

$$8\lambda^3 \geqq 4e_1^3 - 12\lambda^2 e_1.$$

This inequality is equivalent to the inequality

$$(e_1 + \lambda)^2(2\lambda - e_1) \geqq 0.$$

This, however, is impossible, since $e_1 + \lambda > 0$ and $2\lambda - e_1 < 0$. This proves that $g' \neq 0$ for $1 < r_0 < R$.

5. **The sign of** $g'(r_0)$. To show that $g' > 0$ for $1 < r_0 < R$, it is sufficient to show that $g' > 0$ for $r_0 = R^{\frac{1}{2}}$. We have then from (3) that $r = R^{\frac{1}{2}}$. Also $du/d\alpha = g'$ for $r_0 = r = R^{\frac{1}{2}}$. Now $du/d\alpha$ for $\alpha = \omega_1/2$ is equal to $\dfrac{e_1 - 2\lambda}{e_2 - e_3}$, which is positive.

6. **Existence and calculation of** R_2 **and** R_3. Let α and t have the definitions given them in section 4. It is clear from what has been proved that $\lim\,(\omega_3 + t) = \omega_3 + t_0$, say, exists. Hence the maximum of R_2 is e^{t_0}.

Since $2\lambda\alpha + \zeta(\omega_3 + t - \alpha) - \zeta(\omega_3 + t + \alpha) = 0$, we have

$$\lim_{\alpha \to 0}\left(\lambda + \frac{\zeta(\omega_3 + t - \alpha) - \zeta(\omega_3 + t + \alpha)}{2\alpha}\right) = \lambda + \wp(\omega_3 + t_0) = 0\,.$$

This equation has just one solution t_0 such that $0 < t_0 < \omega_1$. Similarly we may show the existence of R_3. Its minimum value is Re^{-t_0}.

7. **Limits of** $g'(r_0)$. It will be sufficient to show that $\lim\limits_{r_0 \to 1} g' = 0$. We have

$$2\lambda\alpha + \zeta(\omega_3 + t - \alpha) - \zeta(\omega_3 + t + \alpha) = 2\lambda\alpha + 2\alpha\wp(\omega_3 + t) + o(\alpha^2) = 0. \tag{11}$$

Hence $\displaystyle\lim_{\alpha \to 0} \frac{2\lambda + 2\wp(\omega_3 + t)}{\alpha} = 0$, t real, subject to (11). Now

$$\frac{du}{d\alpha} = -\frac{2\lambda + 2\wp(\omega_3 + t) + o(\alpha)}{2\alpha\wp'(\omega_3 + t) + o(\alpha^2)}\,.$$

Hence $\displaystyle\lim_{\alpha \to 0} \frac{du}{d\alpha} = 0$. Therefore $\lim\limits_{r_0 \to 1} g' = 0$.

Corresponding results are valid for the region between two concentric spheres.

THE INSTITUTE FOR ADVANCED STUDY.

ON CERTAIN PROPERTIES OF PROJECTIVE PARALLELISM OF SURFACES

By M. L. MacQueen

1. **Introduction.** It is the purpose of this paper to make some contributions to the projective differential geometry of surfaces in ordinary space which are in the nature of projective analogues of certain metric theorems connected with the metric concept of parallelism of surfaces. We provide a projectively defined substitute for the metric property of parallelism of surfaces by employing a projective generalization of euclidean parallelism of surfaces developed in the author's thesis[1] and summarized briefly in a preceding paper.[2] For the basis of our projective theory use is made of one of the well-known transformations of surfaces, namely, the fundamental transformation, or transformation F.[3]

In §2 we introduce a canonical form of the system of differential equations employed in the study of projectively parallel surfaces in ordinary space. Certain results which we shall need for later reference are summarized. In §3, on introducing a projective analogue of the plane at infinity, certain projective properties which are analogues of well-known metric properties of surfaces are investigated. The quadrics of Darboux at corresponding points of two projectively parallel surfaces are considered, and certain configurations connected with these quadrics are studied. A more general type of projective parallelism, which we have called modified projective parallelism of surfaces, is employed in §4 and in the following section. For the study of this type of projective parallelism a somewhat different canonical form of our system of differential equations is used. Finally, projective analogues of certain metric theorems of Bompiani connected with the theory of the correspondence between axial systems of curves on parallel surfaces are obtained in §5.

2. **Analytic basis.** Two surfaces S_x, S_y, in ordinary space S_3, are said to be *projectively parallel*[4] if they are in the relation of a fundamental transformation with the projective normal congruence as the conjugate congruence, and with the developables of the harmonic congruence indeterminate. By considering the projective normal congruence as the common conjugate congruence of the

Received March 25, 1935.

[1] M. L. MacQueen, *A projective generalization of euclidean parallelism of surfaces*, University of Chicago, December, 1933; unpublished doctoral dissertation.

[2] M. L. MacQueen, *A projective generalization of metrically defined associate surfaces*, Transactions of the American Mathematical Society, vol. 36 (1934), p. 826; hereafter referred to as MacQueen, *Associate Surfaces*.

[3] L. P. Eisenhart, *Transformations of Surfaces*, Princeton University Press, 1923, p. 34 et seq.

[4] MacQueen, *Associate Surfaces*, pp. 827–828.

transformation F, we thus provide a projectively defined substitute for the metric normal congruence. Moreover, the assumption that the developables of the harmonic congruence are indeterminate, that is, corresponding tangent planes of the two surfaces intersect in the lines of a fixed plane, affords us a projective substitute for the metric parallelism of the tangent planes, and also a projective analogue of the plane at infinity.

Let us consider two projectively parallel surfaces S_x, S_y with the respective parametric vector equations

$$x = x(u, v)\,, \qquad y = y(u, v)\,.$$

The four coördinates x and the four coördinates y form four pairs of solutions of a completely integrable system of partial differential equations of the form[4]

$$\begin{aligned}
x_{uu} &= px + \alpha x_u + \beta x_v + Ly\,,\\
x_{uv} &= cx + ax_u + bx_v\,,\\
x_{vv} &= qx + \gamma x_u + \delta x_v + Ny\,,\\
y_u &= fx + mx_u + Ay\,,\\
y_v &= gx + nx_v + By\,, \qquad\qquad (mnLN \neq 0)\,.
\end{aligned} \tag{1}$$

The coefficients of the system corresponding to (1) when the rôles of x and y are interchanged will be indicated by dashes and will be given later. In order that S_y may be non-developable we shall assume that $\bar{L}\bar{N} \neq 0$.

System (1) is characterized analytically by the following conditions

$$\begin{aligned}
&\text{(a)}\ \alpha + b + A + (\log N)_u - 3(\log r)_u/2 - 2(\log R)_u = 0\,,\\
&\text{(b)}\ \gamma/r + \alpha + (\log r)_u/2 = 0\,,\\
&\text{(c)}\ \bar{r} = nr/m\,,\\
&\text{(d)}\ f/m = -\,[\log (mn)^{1/2}R/L]_u\,,\\
&\text{(e)}\ m(1-n)\mathfrak{B}'^2 + nr(1-m)\mathfrak{C}'^2 + m_v(\mathfrak{B}' + m_v/4m) + n_u r(\mathfrak{C}' + n_u/4n) = 0\,,
\end{aligned} \tag{2}$$

and by the counterpart of (a), (b), and (d) in the substitution

$$\begin{pmatrix} u & a & c & f & m & p & \alpha & \beta & A & L & r & R \\ v & b & c & g & n & q & \delta & \gamma & B & N & 1/r & rR \end{pmatrix}. \tag{3}$$

A geometrical interpretation of conditions (2) will be found in the paper last cited. The invariants $\mathfrak{B}'$, $\mathfrak{C}'$, R of Green, and the invariant r of Eisenhart, which appear in equations (2), are expressed for the projective lines of curvature on S_x in terms of the coefficients of system (1) by the formulas

$$\begin{aligned}
8\mathfrak{B}' &= 4a + 2N\beta/L - 2\delta + (\log N/L)_v\,,\\
8\mathfrak{C}' &= 4b + 2L\gamma/N - 2\alpha + (\log L/N)_u\,,\\
R &= L\mathfrak{B}'^2/N + \mathfrak{C}'^2\,, \qquad r = N/L\,.
\end{aligned} \tag{4}$$

The coefficients of equations (1) are functions of u, v and satisfy certain integrability conditions[5] which we shall not rewrite here.

We list for future reference the coefficients of the equations corresponding to (1) when the rôles of x and y are interchanged. These coefficients are given by

$$
(5)\qquad
\begin{aligned}
\bar{p} &= A_u + mL - A(m_u/m + f/m + \alpha) - m\beta B/n,\\
\bar{\alpha} &= \alpha + f/m + m_u/m + A, \qquad \bar{\beta} = m\beta/n,\\
\bar{L} &= -m(\alpha f/m + \beta g/n + (f/m)^2 - p - (f/m)_u),\\
\bar{c} &= A_v - A(m_v/m + a) - B(f/n + mb/n),\\
&= B_u - B(n_u/n + b) - A(g/m + na/m),\\
\bar{a} &= a + m_v/m = B + g/m + na/m,\\
\bar{f} &= -A/m, \qquad \bar{m} = 1/m, \qquad \bar{A} = -f/m,
\end{aligned}
$$

and the formulas obtainable therefrom by the substitution (3).

The developables of the projective normal congruence intersect S_x and S_y in the projective lines of curvature which are the parametric curves thereon. The focal points of a projective normal are the points η, ζ defined by

$$
(6)\qquad \eta = y - mx, \qquad \zeta = y - nx.
$$

The curvilinear differential equation defining the asymptotic curves on S_x is

$$
(7)\qquad Ldu^2 + Ndv^2 = 0,
$$

and the asymptotic curves on S_y are given by

$$
(8)\qquad \bar{L}du^2 + \bar{N}dv^2 = 0.
$$

The line of intersection of the tangent planes at corresponding points P_x, P_y of the surfaces S_x, S_y joins the points P_ρ, P_σ defined by

$$
(9)\qquad
\begin{aligned}
\rho &= x_u + fx/m = (y_u - Ay)/m,\\
\sigma &= x_v + gx/n = (y_v - By)/n,
\end{aligned}
$$

as may be seen on inspection of the last two of equations (1). Since the curves corresponding to the developables of the congruence of lines $\rho\sigma$ are indeterminate, the congruence consists entirely of lines lying in a fixed plane. This plane is determined by the two points given by (9) and by the point ρ_u (or σ_v). By differentiating with respect to u the expression for ρ found in (9), and making use of (1), (5), and (9), we may express ρ_u in the form

$$
\rho_u = \bar{L}x/m + Ly + (\alpha + f/m)\rho + \beta\sigma.
$$

[5] Ibid., p. 828.

We therefore see that the fixed plane intersects the projective normal in the point P_τ given by

$$\tau = \bar{L}x + mLy. \tag{10}$$

The same point may be written by symmetry, or by a similar computation using σ_v, in the form

$$\tau = \bar{N}x + nNy. \tag{11}$$

We shall have occasion to use the tetrahedron x, ρ, σ, y as a local tetrahedron of reference with a unit point so chosen that a point

$$x_1 x + x_2 \rho + x_3 \sigma + x_4 y$$

has local coördinates proportional to $x_1, \cdots, x_4$.

3. **The quadrics of Darboux.** The projective generalization of metric parallelism of surfaces summarized in the preceding section will now be employed in formulating projective analogues of certain metric theorems.

Let us consider any point X near a point P_x on surface S_x. The coördinates of such a point X can be represented by power series of the form

$$X = x + x_u \Delta u + x_v \Delta v + (x_{uu}\Delta u^2 + 2x_{uv}\Delta u \Delta v + x_{vv}\Delta v^2)/2 + \cdots,$$

in which $\Delta u, \Delta v$ denote the increments of u, v that correspond to displacement from P_x to the point X. When the derivatives of x of the second and higher orders are replaced by the expressions given for them by (1) and the equations obtained therefrom by differentiation, we find that X can be expressed uniquely in the form

$$X = x_1 x + x_2 x_u + x_3 x_v + x_4 y, \tag{12}$$

where

$$\begin{aligned}
x_1 &= 1 + (p\Delta u^2 + 2c\Delta u \Delta v + q\Delta v^2)/2 + \cdots, \\
x_2 &= \Delta u + (\alpha \Delta u^2 + 2a\Delta u \Delta v + \gamma \Delta v^2)/2 + \cdots, \\
x_3 &= \Delta v + (\beta \Delta u^2 + 2b\Delta u \Delta v + \delta \Delta v^2)/2 + \cdots, \\
x_4 &= (L\Delta u^2 + N\Delta v^2)/2 + [(L_u + \alpha L + AL)\Delta u^3 \\
&\quad + 3aL\Delta u^2 \Delta v + 3bN\Delta u \Delta v^2 + (N_v + \delta N + BN)\Delta v^3]/6 + \cdots.
\end{aligned} \tag{13}$$

These series represent the local coördinates $x_1, \cdots, x_4$ of the point X, referred to the tetrahedron x, x_u, x_v, y with suitably chosen unit point, to terms of as high degree as will be needed. In a similar way, we obtain the power series expansions of the local coördinates of a point Y near P_y on S_y, namely,

$$
(14)\quad
\begin{aligned}
y_1 &= f\Delta u + g\Delta v + [(f_u + fA + mp)\Delta u^2 + 2(af + fm_v/m + bg \\
&\qquad + gn_u/n)\Delta u\Delta v + (g_v + gB + nq)\Delta v^2]/2 + \cdots, \\
y_2 &= m\Delta u + [(\alpha m + f + m_u + mA)\Delta u^2 + 2(mB + g + an)\Delta u\Delta v \\
&\qquad + n\gamma\Delta v^2]/2 + \cdots, \\
y_3 &= n\Delta v + [(m\beta\Delta u^2 + 2(nA + f + bm)\Delta u\Delta v + (\delta n + g + n_v \\
&\qquad + nB)\,\Delta v^2]/2 + \cdots, \\
y_4 &= 1 + A\Delta u + B\Delta v + [(A_u + A^2 + mL)\Delta u^2 + 2(A_v + AB)\Delta u\Delta v \\
&\qquad + (B_v + B^2 + nN)\Delta v^2]/2 + \cdots.
\end{aligned}
$$

The power series expansions of the coördinates of points X, Y, referred to the tetrahedron x, ρ, σ, y, may be easily obtained from (13) and (14). It will be observed that only the first coördinate of each point will be changed. In order to find this coördinate for each point, a simple computation shows that it is sufficient to multiply the series representing the second and third coördinates given in (13) and (14) by $-f/m$ and $-g/n$ respectively, and to add the corresponding results to the series representing the first coördinate in each case. We shall omit the writing of these results.

The equation of any non-singular quadric surface containing the asymptotic tangents of S_x at P_x and having contact of the second order with S_x at P_x can be obtained by writing the equation of a general quadric, and demanding that this equation be satisfied by the series (13) identically in Δu, Δv as far as the terms of the second degree. The result can be written in the form

$$
(15)\qquad Lx_2^2 + Nx_3^2 + 2x_4(-x_1 + k_2x_2 + k_3x_3 + k_4x_4) = 0,
$$

where k_2, k_3, k_4 are arbitrary. Among this three-parameter family of quadric surfaces there is a one-parameter family of quadrics of Darboux, each of which cuts S_x in a curve whose triple point tangents at P_x coincide with the three directions of Darboux[6] for which

$$
(16)\qquad \mathfrak{C}'du^3 - 3\mathfrak{B}'du^2dv - 3r\mathfrak{C}'du\,dv^2 + r\mathfrak{B}'dv^3 = 0.
$$

It is easy to verify that for the quadrics of Darboux at P_x on S_x we have $k_2 = k_3 = 0$; hence these ∞^1 quadrics are represented by the equation

$$
(17)\qquad Lx_2^2 + Nx_3^2 + 2x_4(-x_1 + kx_4) = 0 \qquad (k \text{ arbitrary}).
$$

For the purpose of writing the equation of the quadrics of Darboux referred to the tetrahedron x, ρ, σ, y, a simple computation shows that it is sufficient to replace x_1 in equation (17) by $x_1 + fx_2/m + gx_3/n$. Hence, the equation of any quadric of Darboux at P_x on S_x, referred to the tetrahedron x, ρ, σ, y, is

$$
(18)\qquad Lx_2^2 + Nx_3^2 + 2x_4(-x_1 + k_2x_2 + k_3x_3 + kx_4) = 0,
$$

[6] E. P. Lane, *Bundles and pencils of nets on a surface*, Transactions of the American Mathematical Society, vol. 28 (1926), p. 163.

where k is arbitrary and k_2, k_3 are given by

$$k_2 = -f/m, \qquad k_3 = -g/n. \tag{19}$$

Among the quadrics of Darboux at P_x on S_x there is the quadric of Lie, which is the limit of the quadric determined by three asymptotic tangents of one family of asymptotics constructed at points of a fixed curve of the other family of asymptotics as these points approach coincidence along the fixed asymptotic. It can be shown that the quadric of Lie at P_x on S_x is represented by equation (18), for which the value of k is given by

$$4LNk = pN + qL - LN(m + n). \tag{20}$$

It is evident that the quadric of Darboux at P_x on S_x that passes through P_y on S_y is given by (18) when $k = 0$.

The equation of any quadric of Darboux at P_y on S_y is found in a similar way to be

$$n^2\bar{L}x_2^2 + m^2\bar{N}x_3^2 + 2\, m^2n^2x_1(-x_4 + \bar{k}_2x_2 + \bar{k}_3x_3 + \bar{k}x_1) = 0, \tag{21}$$

where $\bar{k}$ is arbitrary and $\bar{k}_2$, $\bar{k}_3$ are defined by

$$\bar{k}_2 = A/m, \qquad \bar{k}_3 = B/n. \tag{22}$$

We find that the quadric (21) is the quadric of Lie at P_y on S_y in case $\bar{k}$ has the value given by

$$\begin{aligned} 4\bar{L}\bar{N}\bar{k} = A^2\bar{N} + B^2\bar{L} + A\bar{N}(\log A\bar{r}^{1/2})_u + B\bar{L}(\log B/\bar{r}^{1/2})_v + mL\bar{N} \\ + nN\bar{L} - \bar{L}\bar{N}(1/m + 1/n). \end{aligned} \tag{23}$$

The quadric of Darboux at P_y on S_y which passes through P_x on S_x is the quadric (21) for which $\bar{k} = 0$.

Incidentally, it is easy to verify, by use of (2) (a), (3), and the integrability conditions, that (2) (b) may be written in the form

$$A = (\log R/L)_u, \qquad B = (\log R/L)_v. \tag{24}$$

On making use of (24), (19), and (22) we obtain from (2) (d) the relations

$$k_2 - m\bar{k}_2 = (\log mn)_u/2, \qquad k_3 - n\bar{k}_3 = (\log mn)_v/2. \tag{25}$$

In metric geometry we recall that the center of a quadric surface may be defined to be the pole of the plane at infinity. We have remarked earlier in this paper that the fixed plane containing the lines of intersection of the tangent planes at corresponding points of the two projectively parallel surfaces S_x, S_y affords us a projective analogue of the plane at infinity. The points ρ, σ, τ given by (9) and (10) determine this fixed plane, which will be called the *projective plane at infinity*. We shall, therefore, define the *projective center of a quadric as the pole of the projective plane at infinity*.

The first problem that suggests itself is to determine the projective centers of the quadrics of Darboux at corresponding points P_x, P_y of the projectively

parallel surfaces S_x, S_y. The local equations of the polar planes of the points P_ρ, P_σ with respect to the quadrics of Darboux (18) at P_x are, respectively,

$$(26)\qquad Lx_2 - fx_4/m = 0, \qquad Nx_3 - gx_4/n = 0.$$

Therefore, *the line of projective centers of the quadrics of Darboux at a point P_x on the surface S_x joins P_x to the point*

$$(27)\qquad (0, f/mL, g/nN, 1).$$

On finding the point where the polar plane of the point P_r intersects the line defined by (26), we arrive at *the projective center of the quadrics of Darboux at P_x on S_x*, namely,

$$(28)\qquad (-[(f/m)^2/L + (g/n)^2/N + (\bar{L}/mL - 2k)], f/mL, g/nN, 1),$$

the tetrahedron of reference being x, ρ, σ, y. Similarly, the line of projective centers of the quadrics of Darboux at point P_y on the surface S_y is found to join P_y to the point

$$(29)\qquad (1, -mA/\bar{L}, -nB/\bar{N}, 0).$$

The local coördinates of the projective center of the quadrics of Darboux at P_y on S_y are

$$(30)\qquad (1, -mA/\bar{L}, -nB/\bar{N}, -[A^2/\bar{L} + B^2/\bar{N} + mL/\bar{L} - 2\bar{k}]).$$

If two surfaces S and $\bar{S}$ are parallel in the metric sense, a well-known theorem implies that the lines of centers of the quadrics of Darboux at corresponding points of the two surfaces coincide with the normal if, and only if, the two surfaces have constant total curvatures.[7] Inspection of (27) and (29) yields the following projective analogue of the aforementioned metric theorem.

The lines of projective centers of the quadrics of Darboux at corresponding points of two projectively parallel surfaces coincide with the projective normal joining these points if, and only if,

$$(31)\qquad f = g = A = B = 0.$$

The class of projectively parallel surfaces for which these conditions hold would seem to be worthy of some consideration. When $f = g = A = B = 0$, it follows immediately from (19), (22), and (25) that

$$(32)\qquad mn = c \qquad (c = \text{const.}).$$

This condition is reminiscent of a characteristic metric property[8] of the W-surfaces, namely, that a functional relation exists between their principal radii

[7] Fubini and Čech, *Geometria Proiettiva Differenziale*, Bologna, 1926, vol. 1, pp. 177–178.

[8] L. P. Eisenhart, *Differential Geometry*, 1909, p. 291 and p. 296.

of normal curvature. If, further, we make use of (31) and (32), we obtain from (2) (d) the relation

$$R/L = k \qquad (k = \text{const.}).$$

As a result of conditions (32), with the aid of (5), we find from equations (9) that *the reciprocal projective normals of S_x and S_y coincide with the line $\rho\sigma$, and therefore lie in the projective plane at infinity.*[9] Since the curves corresponding to the developables of the reciprocal projective normal congruence have been called[10] the reciprocal projective lines of curvature, we may state the following result.

If the lines of projective centers of the quadrics of Darboux at corresponding points of two projectively parallel surfaces coincide, the reciprocal projective lines of curvature on each of the two surfaces are indeterminate.

From (28) and (30) it can be seen that the projective center of the quadric of Darboux at P_x on S_x which passes through P_y coincides with the projective center of the quadric of Darboux at P_y on S_y which passes through P_x. The common projective center of these quadrics is the point $(\bar{L}, 0, 0, -mL)$. On referring to (10) we may state the following conclusion.

When the lines of projective centers of the quadrics of Darboux at corresponding points of two projectively parallel surfaces coincide, the projective centers of the quadrics of Darboux at P_x, P_y which pass through P_y, P_x, respectively, coincide in a point on the projective normal. The corresponding points P_x, P_y are separated harmonically by this point and the point where the projective normal is intersected by the projective plane at infinity.

Returning now to the situation where no restriction is imposed on the position of the lines of projective centers of the quadrics of Darboux at corresponding points of two projectively parallel surfaces, we shall investigate the developables of the congruences generated by these two lines. Since, from (27), the line of projective centers of the quadrics of Darboux at P_x on S_x joins P_x to the point

$$\phi = f\rho/mL + g\sigma/nN + y, \tag{33}$$

it is evident that the point ξ defined by

$$\xi = \phi + \mu x \qquad (\mu \text{ scalar}) \tag{34}$$

is any point, except P_x, on this line. By the usual method we obtain the differential equation of the developables of the line of projective centers of the quadrics of Darboux at P_x on S_x, namely,

$$\begin{aligned} &(f\beta r/m + g\gamma/nr - fg/mn + (g/n)_u)du^2 \\ &\quad + (fr[f/m + 2b - (\log fr^{\frac{1}{2}}/m)_u]/m - g[g/n + 2a - (\log g/nr^{\frac{1}{2}})_v]/n \\ &\quad - N(m - n))dudv - r(f\beta r/m + g\gamma/nr - fg/mn + (f/m)_v)dv^2 = 0. \end{aligned} \tag{35}$$

[9] Fubini and Čech, loc. cit., pp. 177–178.

[10] E. P. Lane, *Contributions to the theory of conjugate nets*, American Journal of Mathematics, vol. 49 (1927), p. 570.

The equation for the determination of the focal points of the line of projective centers of the quadrics of Darboux at P_x is

$$\mu^2 + S\mu + T = 0, \tag{36}$$

where we have placed

$$S = f[f/m + (\log fr^{\frac{1}{2}}/m)_u]/mL + g[g/n + (\log g/nr^{\frac{1}{2}})_v]/nN + m + n, \tag{37}$$

and where the expression for T is not needed in what follows. If μ_1, μ_2 are the roots of equation (36), then the corresponding points ξ_1, ξ_2 given by the formula (34) are the focal points.

Since the line of projective centers of the quadrics of Darboux at P_y on S_y joins P_y to the point P_ψ defined by

$$\psi = - mA\rho/\bar{L} - nB\sigma/\bar{N} + x, \tag{38}$$

the point

$$\zeta = \psi + \nu y \qquad (\nu \text{ scalar}) \tag{39}$$

is any point, except P_y, on this line. By use of (39), we obtain, similarly, the differential equation of the developables of the line of projective centers of the quadrics of Darboux at a point P_y on the surface S_y, namely,

$$\begin{aligned} &m(A\beta r + B\gamma/r + AB + B_u)du^2 \\ &+ (Anr[A - 2b + (\log A(r/mn^3)^{\frac{1}{2}})_u] - mB[B - 2a + (\log B/(m^3nr)^{\frac{1}{2}})_v] \\ &- m\bar{N}(1/m - 1/n))dudv - nr(A\beta r + B\gamma/r + AB + A_v)dv^2 = 0. \end{aligned} \tag{40}$$

If ν_1, ν_2 are the roots of

$$\nu^2 + \bar{S}\nu + \bar{T} = 0, \tag{41}$$

in which $\bar{S}$ is defined by the formula

$$\bar{S} = A[A - (\log A(nr/m)^{\frac{1}{2}})_u]/\bar{L} + B[B - (\log B(m/nr)^{\frac{1}{2}})_v]/\bar{N} + 1/m + 1/n, \tag{42}$$

the definition of $\bar{T}$ not being essential to our work, the corresponding points ζ_1, ζ_2, given by (39), are the focal points of the line of projective centers of the quadrics of Darboux at P_y on S_y.

Demoulin has shown,[11] in the metric theory, that the developables of the congruence generated by the line of centers of the quadrics of Darboux at a point of a surface intersect the surface in a conjugate net. The harmonic invariant of (7) and (35) vanishes if, and only if, $(f/m)_v = (g/n)_u$, a condition which is easily seen to be satisfied by inspection of (2) (d) and (3). Similarly, the harmonic invariant of (8) and (40) is found to vanish if, and only if, $A_v =$

[11] Demoulin, *Sur quelques propriétés des surfaces courbes*, Comptes Rendus, vol. 147 (1908), p. 565.

B_u. Reference to the integrability conditions[12] shows that this condition is satisfied. Thus the following theorem is established.

The developables of the congruences generated by the lines of projective centers of the quadrics of Darboux at corresponding points of two projectively parallel surfaces intersect the respective surfaces in conjugate nets.

The projective center of the quadric of Lie at P_x on S_x is given by (28), if the value of k, defined by (20), is substituted therein. Making this substitution, we find that the projective center of this quadric is the point

$$\xi = \phi + \mu x$$

wherein ϕ is defined by (33) and

$$(43) \quad \mu = -[(f/m)^2/L + (g/n)^2/N + \bar{L}/mL - (p/L + q/N - m - n)/2].$$

The cross ratio of point P_x, the projective center of the quadric of Lie at P_x, and the focal points of the line of projective centers of the quadrics of Darboux at P_x is given by

$$(\infty,\ \mu,\ \mu_1,\ \mu_2),$$

where μ is defined by (43) and μ_1, μ_2 are the roots of (36). This cross ratio is harmonic if, and only if,

$$\mu_1 + \mu_2 = 2\mu.$$

On making use of (36) and (43), together with (2) (b), (2) (c), (3), and (5), we find that this condition is satisfied. We therefore arrive at a projective generalization of a theorem[13] of Demoulin. A similar result can be obtained by considering the quadric of Lie at P_y on S_y. Combining these results we have the theorem

The two focal points of the line of projective centers of the quadrics of Darboux at $P_x(P_y)$ on $S_x(S_y)$ regarded as generating a congruence separate harmonically the point $P_x(P_y)$ and the projective center of the quadric of Lie at $P_x(P_y)$.

We interpolate here a few remarks on the axes at corresponding points of two projectively parallel surfaces. The axis of the projective lines of curvature at a point P_x of the surface S_x is the line of intersection of the osculating planes of the projective lines of curvature through P_x. By making use of equations (1) it is not difficult to show that the axis at P_x joins P_x to the point $(0, \gamma/N, \beta/L, 1)$. Similarly, by use of (5) and the equations corresponding to (1) when the rôles of x and y are interchanged, it is found that the axis at P_y joins P_y to the point $(1, \gamma n/\bar{N}, \beta m/\bar{L}, 0)$. It can be shown that the axis at P_x intersects the fixed plane determined by the points ρ, σ, τ in the point $(\bar{L}/mL, \gamma/N, \beta/L, 1)$. Furthermore, we find that the axis at P_y intersects this fixed plane in the same point. Thus we have established a projective analogue of the property of the parallelism of the axes in the metric theory of parallel surfaces. Our result may therefore be stated as follows.

[12] MacQueen, *Associate Surfaces*, p. 829.

[13] Demoulin, loc. cit., p. 565.

The axes of the projective lines of curvature at corresponding points of two projectively parallel surfaces intersect in a point which lies in the projective plane at infinity.

4. **Modified projective parallelism of surfaces.** In this and in the following section we shall drop the assumption that the common conjugate congruence of the transformation F is the projective normal congruence, and shall employ in its place a general conjugate congruence. The configuration composed of two surfaces in ordinary space in the relation of a fundamental transformation having a general conjugate congruence and with the developables of the harmonic congruence indeterminate leads us to a characterization of surfaces which are projectively parallel in a modified[14] sense.

For the analytic basis of our study of modified projective parallelism a somewhat different canonical form of the basic system of differential equations is employed. If S_x, S_y are a pair of surfaces projectively parallel in the modified sense, the four coördinates x and the four coördinates y form four pairs of solutions of a completely integrable system of differential equations[14] of the form

$$\begin{aligned} x_{uu} &= L(x+y) + \alpha x_u + \beta x_v, \\ x_{uv} &= a x_u + b x_v, \\ x_{vv} &= N(x+y) + \gamma x_u + \delta x_v, \\ y_u &= m x_u, \qquad y_v = n x_v, \qquad (mnLN \neq 0). \end{aligned} \tag{44}$$

The integrability conditions for this system are found to be

$$\begin{aligned} & a_u + ab = \alpha_v + \beta\gamma, \qquad b_v + ab = \delta_u + \beta\gamma, \\ & b_u + b^2 + a\beta = \beta_v + b\alpha + nL + \beta\delta + L, \\ & a_v + a^2 + b\gamma = \gamma_u + a\delta + mN + \alpha\gamma + N, \\ & L_v = aL - \beta N, \qquad N_u = bN - \gamma L, \\ & m_v = a(n-m), \qquad n_u = b(m-n). \end{aligned} \tag{45}$$

The coefficients of the equations corresponding to (44), when the rôles of x and y are interchanged, are indicated by dashes and are given by the following expressions:

$$\begin{aligned} & \bar{L} = mL, \qquad \bar{\alpha} = \alpha + m_u/m, \qquad \bar{\beta} = m\beta/n, \\ & \bar{a} = na/m, \qquad \bar{b} = mb/n, \qquad \bar{m} = 1/m,\ \bar{n} = 1/n, \\ & \bar{N} = nN, \qquad \bar{\gamma} = n\gamma/m, \qquad \bar{\delta} = \delta + n_v/n. \end{aligned} \tag{46}$$

We shall assume that $\bar{L}\bar{N} \neq 0$ in order that S_y may be non-developable.

[14] MacQueen, loc. cit., p. 832.

The line of intersection of the tangent planes at corresponding points of the two surfaces joins the points ρ, σ defined by

$$\rho = x_u, \qquad \sigma = x_v. \tag{47}$$

It is easily shown that the fixed plane containing the lines $\rho\sigma$ crosses the line xy at the point

$$\tau = x + y. \tag{48}$$

The differential equation of the asymptotic curves on S_x is

$$L du^2 + N dv^2 = 0, \tag{49}$$

and the asymptotic curves on S_y are given by

$$mL du^2 + nN dv^2 = 0. \tag{50}$$

Some of the invariants of the parametric conjugate net N_x are found to have in our notation the following formulas:

$$\begin{aligned} 8\mathfrak{B}' &= 6a - 2\delta - 3(\log L)_v + (\log N)_v, \\ 8\mathfrak{C}' &= 6b - 2\alpha - 3(\log N)_u + (\log L)_u, \\ H &= ab - a_u, \qquad K = ab - b_v. \end{aligned} \tag{51}$$

By use of (45) and (46), the corresponding invariants for N_y, indicated by dashes, are given by the following expressions:

$$\begin{aligned} 8\bar{\mathfrak{B}}' &= 8\mathfrak{B}' + (\log m^3/n)_v, \qquad & 8\bar{\mathfrak{C}}' &= 8\mathfrak{C}' + (\log n^3/m)_u, \\ \bar{H} &= H - (\log m)_{uv}, \qquad & \bar{K} &= K - (\log n)_{uv}. \end{aligned} \tag{52}$$

Several results similar to those obtained in §3 will now be given. Inasmuch as the proofs of these results run parallel to those in the section just cited, they will be omitted here. The equation of any quadric of Darboux at a point P_x on the surface S_x, referred to the tetrahedron x, x_u, x_v, y, is found to be

$$Lx_2^2 + Nx_3^2 + 2x_4(-x_1 + k_2x_2 + k_3x_3 + kx_4) = 0, \tag{53}$$

where k is arbitrary and k_2, k_3 are given by

$$k_2 = 2\mathfrak{C}' - b - \gamma/r, \qquad k_3 = 2\mathfrak{B}' - a - \beta r. \tag{54}$$

Similarly, the equation of any quadric of Darboux at point P_y on surface S_y, referred to the same tetrahedron, is

$$nLx_2^2 + mNx_3^2 + 2mnx_1(-x_4 + \bar{k}_2x_2 + \bar{k}_3x_3 + \bar{k}x_1) = 0, \tag{55}$$

where $\bar{k}$ is arbitrary and $\bar{k}_2$, $\bar{k}_3$ are expressed by the formulas

$$m\bar{k}_2 = k_2 - (\log mn)_u/4, \qquad n\bar{k}_3 = k_3 - (\log mn)_v/4. \tag{56}$$

The line of projective centers of the quadrics of Darboux at P_x on S_x joins P_x to the point ϕ given by

$$\phi = -k_2x_u/L - k_3x_v/N + y, \tag{57}$$

and the projective center is the point with coördinates

$$(-[k_2^2/L + k_3^2/N + 1 - 2k], -k_2/L, -k_3/N, 1). \tag{58}$$

Similarly, the line of projective centers of the quadrics of Darboux at P_y on S_y is found to join P_y to the point ψ defined by

$$\psi = -m\bar{k}_2x_u/L - n\bar{k}_3x_v/N + x, \tag{59}$$

and the projective center has local coördinates

$$(1, -m\bar{k}_2/L, -n\bar{k}_3/N, -[m\bar{k}_2^2/L + n\bar{k}_3^2/N + 1 - 2\bar{k}]). \tag{60}$$

Inspection of (56), (57), and (59) leads us to the following theorem.

If two surfaces S_x, S_y are projectively parallel in the modified sense, and if the line of projective centers of the quadrics of Darboux at a point P_x on the surface S_x coincides with the line xy joining corresponding points of the two surfaces, the line of projective centers of the quadrics of Darboux at a point P_y on the surface S_y will also coincide with the line xy if, and only if, $mn =$ const.

The reciprocal line of the line xy with respect to a quadric of Darboux associated with the point P_x of the surface S_x intersects the tangents of the parametric curves at P_x in the points

$$s = x_u + k_2x, \qquad t = x_v + k_3x. \tag{61}$$

Similarly, the reciprocal line of the line xy with respect to a quadric of Darboux associated with the corresponding point P_y of the surface S_y crosses the parametric tangents through P_y in the points

$$\bar{s} = y_u + m\bar{k}_2y, \qquad \bar{t} = y_v + n\bar{k}_3y. \tag{62}$$

When the lines of projective centers of the quadrics of Darboux at corresponding points P_x, P_y of the two surfaces coincide, equations (61) and (62) show, with the aid of (44), that *each of the reciprocal lines of the line xy with respect to a quadric of Darboux associated with the corresponding points P_x, P_y of the two surfaces S_x, S_y coincides with the line $\rho\sigma$, and therefore lies in the projective plane at infinity.* It follows that the curves on the two surfaces corresponding to the developables of the congruences generated by the reciprocal lines are indeterminate.

The developables of the congruences generated by the lines of projective centers of the quadrics of Darboux at corresponding points of the surfaces S_x, S_y can be found by the usual method. On considering equation (34) with ϕ de-

fined by (67), the differential equation of the developables of the line of projective centers of the quadrics of Darboux at P_x on S_x is found to be

$$\begin{aligned}(63)\qquad &(k_2k_3 + \beta rk_2 + \gamma k_3/r + k_{3u})du^2 + (rk_2[k_2 + 2b - (\log k_2r^{1/2})_u]\\ &- k_3[k_3 + 2a - (\log k_3/r^{1/2})_v] + N(m-n))\,dudv\\ &- (k_2k_3 + \beta rk_2 + \gamma k_3/r + k_{2v})\,dv^2 = 0.\end{aligned}$$

The focal points of the line of projective centers of these quadrics are defined by

$$(64)\qquad \xi_{1,2} = \phi + \mu_{1,2}x,$$

if $\mu_{1,2}$ are the roots of an equation of the form

$$(65)\qquad \mu^2 + S\mu + T = 0,$$

where S is defined by placing

$$(66)\qquad S = k_2^2/L + k_3^2/N - k_2(\log k_2r^{1/2})_u/L - k_3(\log k_3/r^{1/2})_v/N + m + n,$$

and where the expression for T is not needed in what follows. Similarly, the differential equation of the curves on S_y corresponding to the developables of the congruence generated by the line of projective centers of the quadrics of Darboux at P_y is given by

$$\begin{aligned}(67)\qquad &m(\beta mr\bar{k}_2 + \gamma n\bar{k}_3/r + (n\bar{k}_3)_u + mn\bar{k}_2\bar{k}_3)\,du^2\\ &+ (mnr\bar{k}_2[m\bar{k}_2 + 2b - (\log \bar{k}_2(mr/n^3)^{1/2})_u]\\ &- mn\bar{k}_3[n\bar{k}_3 + 2a - (\log \bar{k}_3(n/rm^3)^{1/2})_v] - N(m-n))\,dudv\\ &- nr(\beta mr\bar{k}_2 + \gamma n\bar{k}_3/r + (m\bar{k}_2)_v + mn\bar{k}_2\bar{k}_3)\,dv^2 = 0.\end{aligned}$$

When ψ is defined by (59), the focal points of the line of projective centers of the quadrics at P_y are

$$(68)\qquad \zeta_{1,2} = \psi + \nu_{1,2}y,$$

if $\nu_{1,2}$ are the roots of

$$(69)\qquad \nu^2 + \bar{S}\nu + \bar{T} = 0,$$

and if

$$\begin{aligned}(70)\qquad \bar{S} = m\bar{k}_2^2/L + n\bar{k}_3^2/N - \bar{k}_2(\log \bar{k}_2(mnr)^{1/2})_u/L&\\ - \bar{k}_3(\log \bar{k}_3(mn/r)^{1/2})_v/N + 1/m + 1/n,&\end{aligned}$$

the expression for $\bar{T}$ being omitted.

Calculation of the harmonic invariant of (49) and (63) shows that the curves defined by (63) form a conjugate net on S_x if, and only if,

$$(71)\qquad k_{2v} = k_{3u},$$

a condition which is easily shown to be satisfied by differentiating (54) and making use of (51) and the integrability conditions. The harmonic invariant of (50) and (67) vanishes if, and only if,

$$(72) \qquad (m\bar{k}_2)_v = (n\bar{k}_3)_u.$$

On referring to (56) it becomes evident that this condition is equivalent to (71). Combining these results we see that *the lines of projective centers of the quadrics of Darboux at corresponding points of two modified projectively parallel surfaces each generate congruences whose developables intersect the respective surfaces in conjugate nets.*

The quadric (53) is the quadric of Lie at P_x on S_x in case k_2, k_3 have the values given in equation (54) and k has the value given by

$$(73) \qquad \begin{aligned} 4LNk = Nk_2^2 + Lk_3^2 + Nk_2(\log k_2 r^{1/2})_u + Lk_3(\log k_3/r^{1/2})_v \\ + LN(2 - m - n). \end{aligned}$$

Similarly, we find that the quadric (55) is the quadric of Lie at P_y on S_y in case $\bar{k}_2$, $\bar{k}_3$ are defined by (56) and $\bar{k}$ has the value

$$(74) \qquad \begin{aligned} 4LN\bar{k} = mN\bar{k}_2^2 + nL\bar{k}_3^2 + N\bar{k}_2 (\log \bar{k}_2(mnr)^{1/2})_u + L\bar{k}_3(\log \bar{k}_3(mn/r)^{1/2})_v \\ + LN\,(2 - 1/m - 1/n). \end{aligned}$$

Substituting in (58) the value of k given in (73), we find that the projective center of the quadric of Lie at P_x on S_x is the point

$$\xi = \phi + \mu x,$$

where ϕ is given by (57) and

$$\mu = -\,[k_2^2/L + k_3^2/N - k_2(\log k_2 r^{1/2})_u/L - k_3(\log k_3/r^{1/2})_v/N + m + n]/2.$$

On considering the congruence generated by the line of projective centers of the quadrics of Darboux at P_x on S_x, it may be shown that *the foci of the generator separate harmonically the point P_x and the projective center of the quadric of Lie at P_x.* A similar result may be obtained at point P_y on surface S_y.

5. **Axial systems.** A curve on a surface is called a union curve of a congruence if the curve is such that the osculating plane at each of its points contains the line of the congruence which passes through the point but which does not lie in the tangent plane of the surface at the point. The union curves of a congruence have also been called an axial system of curves, since the osculating planes at a point of the surface of the ∞^1 curves that pass through this point form a pencil with the line of the congruence as axis.

In the metric theory of surfaces Bompiani has shown[15] that there always exists a system of axial curves on a surface S to which corresponds a system of

[15] E. Bompiani, *Corrispondenza fra una superficie e le sue parallele*, Mathematische Zeitschrift, vol. 24 (1925–26), p. 311 et seq.

axial curves on a surface $\bar{S}$ parallel to S. The axis of the pencil which is thus determined at corresponding points P, $\bar{P}$ of each of the two parallel surfaces S, $\bar{S}$ is called *the axis of the correspondence.* Moreover, Bompiani has shown that the locus of the axis of the correspondence at a point P of the surface S, as the surface $\bar{S}$ varies, is a quadric cone, with vertex at the point P, which contains the tangents of the lines of curvature through P. This cone is found to be composite only when the lines of curvature on S are spherical or plane. The cone contains the normal to S at P if, and only if, the axial curves on S are the geodesics that pass through each point of S. A characteristic property of minimal surfaces among the W-surfaces is the possession of geodesic curves as axial curves.

The present section is devoted to the problem of establishing projective analogues of the metric theorems stated above. We begin by choosing a point P_z on the line xy joining corresponding points P_x, P_y of the surfaces S_x, S_y, in §4, such that

$$y = (1 + k)z + kx \qquad (k = \text{const.}). \tag{75}$$

If we make the transformation (75) on system (44), we find that x, z are solutions of differential equations of the form

$$\begin{aligned} x_{uu} &= L'(x + z) + \alpha' x_u + \beta' x_v, \\ x_{uv} &= a' x_u + b' x_v, \\ x_{vv} &= N'(x + z) + \gamma' x_u + \delta' x_v, \\ z_u &= m' x_u, \qquad z_v = n' x_v, \end{aligned} \tag{76}$$

wherein the coefficients, indicated by primes, are given by the formulas

$$\begin{aligned} &L' = (1 + k)L, \qquad \alpha' = \alpha, \qquad \beta' = \beta, \\ &a' = a, \quad b' = b, \quad m' = (m - k)/(1 + k), \quad n' = (n - k)/(1 + k), \\ &N' = (1 + k)N, \qquad \gamma' = \gamma, \qquad \delta' = \delta. \end{aligned} \tag{77}$$

The integrability conditions (45) are satisfied identically by these coefficients. The coefficients of the equations corresponding to (76) when the rôles of x and z are interchanged may be found by use of (46) and (77). If the fourth equation of (76) is differentiated with respect to v and use is made of the second, fourth, and fifth of equations (76), we obtain a Laplace equation from which it follows that the parametric net N_z is conjugate. It can easily be shown that the tangent planes at corresponding points of the two surfaces S_x, S_z intersect in the lines of a fixed plane. Furthermore, it can be seen that the tangent planes at corresponding points P_x, P_y, P_z of the surfaces S_x, S_y, S_z intersect in the same line by differentiating equation (75) to obtain

$$y_u = (1 + k)z_u + kx_u, \qquad y_v = (1 + k)z_v + kx_v.$$

In order that the surfaces S_x, S_y, S_z be distinct, we exclude the cases for which $k = 0$ and $k = -1$.

Let us consider a curve C_λ of the family $dv - \lambda du = 0$ through point P_x on surface S_x. The osculating plane of C_λ at P_x is determined by the points x, x', x'', where

$$x' = x_u + x_v\lambda, \qquad x'' = x_{uu} + 2x_{uv}\lambda + x_{vv}\lambda^2 + x_v\lambda' \qquad (\lambda' = \lambda_u + \lambda\lambda_v).$$

The coördinates ξ of the osculating plane of C_λ at P_x, referred to the tetrahedron x, ρ, σ, y, are found by a simple calculation to be given by

$$\begin{aligned} &\xi_1 = 0, \qquad \xi_2 = -\lambda(L + N\lambda^2), \qquad \xi_3 = L + N\lambda^2, \\ &\xi_4 = -\lambda' - \beta - (2b - \alpha)\lambda + (2a - \delta)\lambda^2 + \gamma\lambda^3. \end{aligned} \tag{78}$$

Similarly, the coördinates η of the osculating plane of the corresponding curve through point P_z on S_z, referred to the same tetrahedron, are found to be

$$\begin{aligned} &\eta_1 = P, \qquad \eta_2 = \lambda(1 + k)(n - k)Q, \\ &\eta_3 = -(1 + k)(m - k)Q, \qquad \eta_4 = kP, \end{aligned} \tag{79}$$

where we have placed

$$\begin{aligned} P &= (m - k)L + (n - k)N\lambda^2, \\ Q &= (m - k)(n - k)\lambda' + \beta(m - k)^2 + [2b(m - k)^2 - \alpha(m - k)(n - k) \\ &\quad - m_u(n - k)]\,\lambda - [2a(n - k)^2 - \delta(m - k)(n - k) \\ &\quad - n_v(m - k)]\,\lambda^2 - \gamma(n - k)^2\lambda^3. \end{aligned} \tag{80}$$

The differential equation of the axial curves on the surface S_x of the congruence generated by the line joining P_x to the point P_h, with coördinates $(0, h_2, h_3, 1)$, can be written by means of the condition of united position $h_2\xi_2 + h_3\xi_3 + \xi_4 = 0$ and equations (78). We find, after replacing λ by v', the axial curves on S_x to be given by

$$v'' = Lh_3 - \beta - (Lh_2 + 2b - \alpha)v' + (Nh_3 + 2a - \delta)v'^2 - (Nh_2 - \gamma)v'^3. \tag{81}$$

In a similar way we obtain, on the surface S_z, the differential equation of the axial curves of the congruence generated by the line joining P_z to the point $(1, k_2, k_3, 0)$, namely,

$$\begin{aligned} v'' &= (1 + k)(m - k)Lk_3/(n - k) - \beta(m - k)/(n - k) \\ &\quad - [(1 + k)Lk_2 + 2b(m - k)/(n - k) - \alpha - m_u/(m - k)]\,v' \\ &\quad + [(1 + k)Nk_3 + 2a(n - k)/(m - k) - \delta - n_v/(n - k)]v'^2 \\ &\quad - [(1 + k)(n - k)Nk_2/(m - k) - \gamma(n - k)/(m - k)]\,v'^3. \end{aligned} \tag{82}$$

Now if to every curve (81) on S_x there corresponds a curve (82) on S_z, and vice versa, we find necessary and sufficient conditions for this correspondence to be

$$
(83)\qquad
\begin{aligned}
Lh_3 &= (1 + k)(m - k)Lk_3/(n - k) + \beta(n - m)/(n - k),\\
Lh_2 &= (1 + k)Lk_2 - m_u/(m - k) - 2b(n - m)/(n - k),\\
Nh_3 &= (1 + k)Nk_3 - n_v/(n - k) - 2a(m - n)/(m - k),\\
Nh_2 &= (1 + k)(n - k)Nk_2/(m - k) + \gamma(m - n)/(m - k).
\end{aligned}
$$

From these equations we obtain

$$(84)\qquad h_2 = (m_1 + m_u t/(m - n))/L, \qquad h_3 = (m_2 + n_v/t(n - m))/N,$$

wherein we have placed

$$m_1 = \gamma/r - 2b, \qquad m_2 = \beta r - 2a, \qquad t = (n - k)/(m - k).$$

Let the axis of the correspondence joining P_x to the point P_h be represented by the equations

$$(85)\qquad x_2 - h_2 x_4 = 0, \qquad x_3 - h_3 x_4 = 0,$$

with h_2, h_3 given by (84). The locus of this line, as the surface S_x varies, is found by eliminating t from equations (85). For this result we obtain *the equation of a quadric cone with its vertex at the point P_x, namely,*

$$(86)\qquad LNx_2x_3 - (m_1Lx_2 + m_2Nx_3)x_4 + [m_1m_2 + m_un_v/(m - n)^2]\,x_4^2 = 0.$$

It is now possible to deduce several interesting theorems. For example, *the cone* (86) *intersects the tangent plane, $x_4 = 0$, at point P_x of the surface S_x in the tangents of the parametric conjugate net through P_x.* It is easily shown that *the cone* (86) *is composite if, and only if,*

$$(87)\qquad m_un_v = 0.$$

If $m_u = 0$, by use of (45) and (51), it follows that $H = 0$ when $m \neq n$. Then *the u-curves on S_x are cone curves,* namely, the curves of contact of a cone circumscribing the surface. Similarly, if $n_v = 0$, it follows that $K = 0$, and the v-curves are cone curves. For the net N_x the invariants corresponding to H, K are found to have the following expressions:

$$(88)\qquad \bar{H}' = \frac{n - k}{m - k}\left(H + \frac{am_u}{m - k}\right), \qquad \bar{K}' = \frac{m - k}{n - k}\left(K + \frac{bn_v}{n - k}\right).$$

In the presence of conditions (87), inspection of (52) and (88) shows that the parametric curves are cone curves on S_x and on every surface projectively parallel to S_x in the modified sense. *The cone* (86) *contains the line xy if, and only if,*

$$m_1m_2 + m_un_v/(m - n)^2 = 0,$$

and in this case the axial curves on S_x are what we call *the modified projective geodesics* on S_x, i.e., the union curves of the congruence of lines xy.

SOUTHWESTERN COLLEGE.

A FUNCTION NOT CONSTANT ON A CONNECTED SET OF CRITICAL POINTS

By Hassler Whitney

1. **Introduction.** Let $f(x_1, \cdots, x_n)$ be a function of class C^m (i.e., with continuous partial derivatives through the mth order) in a region R. Any point at which all its first partial derivatives vanish is called a *critical point* of f. Suppose every point of a connected set A of points in R is a critical point. It is natural to suspect then that f is a constant on A. But this need not be so. We construct below an example with $n = 2$, $m = 1$, $A =$ an arc. The example may be extended to the case $n = n$, $m = n - 1$, $A =$ an arc. The arc and the function on the arc are easily defined. The extension of the function through the rest of the plane or space is given by a theorem of the author.[1]

The question settled in this paper was raised implicitly in a paper of W. M. Whyburn.[2] It is brought up by his definition of critical sets as the maximal connected subsets of the set of critical points on which the function takes a single critical value. Theorem 2 of Whyburn's paper shows that an example of the type given in the present paper can be constructed only by using critical sets which have points that cannot be joined in these sets by rectifiable arcs. It would be interesting to discover how far from rectifiable a closed set must be to be a set of critical points of some function but not a critical set of the function. It may be remarked that any closed set may be a critical set.[3]

For fixed n and m large enough, $m \geqq [(n - 3)^2/16 + n]$, where $[n]$ is the integral part of n, f must be constant on any connected critical set, as shown by M. Morse and A. Sard in an unpublished paper.

The example shows that it is in general impossible to express the values of a function $f(x_1, \cdots, x_n)$ along a curve which is not rectifiable by means of an integral of a function of partial derivatives of f of order $\leqq n - 1$ along the curve.[4]

2. **The arc.** Let Q be a square of side 1 in the plane. Let Q_0, Q_1, Q_2, Q_3 be squares of side 1/3 lying interior to Q in cyclical order, each a distance 1/12

Received May 16, 1935; presented to the American Mathematical Society, October 26, 1935. The example given here was discovered in 1932 while the author was a National Research Fellow.

[1] H. Whitney, *Analytic extensions of differentiable functions defined in closed sets*, Transactions of the American Mathematical Society, vol. 36 (1934), pp. 63–89, Lemma 2. We refer to this paper as AE.

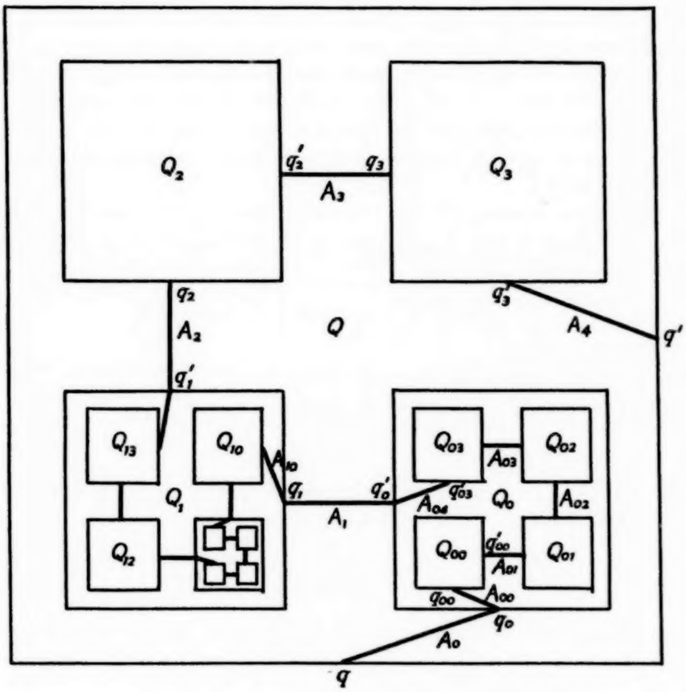
[2] W. M. Whyburn, Bull. Amer. Math. Soc., vol. 35 (1929), pp. 701–708.

[3] See A. Ostrowski, Bull. des Sciences Math., Feb. (1934), pp. 64–72.

[4] For such an expression (using partial derivatives of any desired order) along a rectifiable curve, see H. Whitney, *Functions differentiable on the boundaries of regions*, Annals of Mathematics, vol. 35 (1934), pp. 482–485, (1) and (3).

from the boundary of Q. Let q and q' be the centers of the sides of Q along Q_0, Q_1, and along Q_3, Q_0. Let q_i and q'_i be centers of adjacent sides of Q_i (i = 0, 1, 2, 3) so that q'_{i-1} and q_i face each other ($i = 1, 2, 3$), and q_0 is near q, q'_3 is near q'. Let A_0 be a line joining q and q_0, let A_i join q'_{i-1} and q_i ($i = 1, 2, 3$), and let A_4 join q'_3 and q'.

Suppose we have constructed squares $Q_{i_1\cdots i_t}$, points $q_{i_1\cdots i_t}$, $q'_{i_1\cdots i_t}$, and lines $A_{j_1\cdots j_t}$ (each $i_k = 0, 1, 2, 3$; each $j_k = 0, 1, 2, 3, 4$) for $t < s$. By taking a square $Q_{i_1\cdots i_{s-2}}$, shrinking it to a third its size, and turning it around and upside down if necessary, we may place it in $Q_{i_1\cdots i_{s-1}}$ so that $q_{i_1\cdots i_{s-2}}$ and $q'_{i_1\cdots i_{s-2}}$ go into $q_{i_1\cdots i_{s-1}}$



and $q'_{i_1\cdots i_{s-1}}$, and thus construct squares $Q_{i_1\cdots i_s}$, etc. We continue this process indefinitely. Let $Q_{i_1 i_2\cdots}$ be the point common to Q, Q_{i_1}, $Q_{i_1 i_2}$, $\cdots$ for each $(i_1, i_2, \cdots)$.

The line segments $A_{i_1\cdots i_s}$ together with the points $Q_{i_1 i_2\cdots}$ form an arc A. It may be represented as the topological image of the segment (0, 1) by letting $A_{i_1\cdots i_s}$ correspond to the segment

$$\left(\frac{2i_1+1}{9} + \cdots + \frac{2i_{s-1}+1}{9^{s-1}} + \frac{2i_s}{9^s},\ \frac{2i_1+1}{9} + \cdots + \frac{2i_{s-1}+1}{9^{s-1}} + \frac{2i_s+1}{9^s}\right),$$

and letting $Q_{i_1 i_2 \cdots}$ correspond to the number

$$\frac{2i_1 + 1}{9} + \frac{2i_2 + 1}{9^2} + \cdots .$$

3. **The function** $f(x, y)$. We first define $f(x, y)$ along the arc A as follows:

$$\text{on } A_{i_1 \cdots i_s}, \qquad f = \frac{i_1}{4} + \cdots + \frac{i_s}{4^s};$$

$$\text{at } Q_{i_1 i_2 \cdots}, \qquad f = \frac{i_1}{4} + \frac{i_2}{4^2} + \cdots .$$

f increases from 0 to 1 as we run along A from q to q'. Set $f_{00}(x,y) = f(x,y)$, $f_{10}(x,y) = f_{01}(x,y) = 0$ on A. We shall show that *f_{00} is of class C^1 on A in terms of* (f_{00}, f_{10}, f_{01}) (see AE). It will follow from Lemma 2 of AE that the definition of $f(x,y)$ may be extended over the plane (in particular, over Q) so that f is of class C^1; also $\partial f/\partial x = f_{10} = 0$, $\partial f/\partial y = f_{01} = 0$ on A, and hence each point of A is a critical point of f.

As f_{10} and f_{01} are continuous in A, we need merely prove that for each $\epsilon > 0$ there is a $\delta > 0$ such that if (x,y) and (x',y') are points of A whose distance apart is $r < \delta$, then

$$|f(x',y') - f(x,y)| < r\epsilon, \tag{1}$$

(see AE (3.1) and (3.2)).

The proof rests on the following two facts.

(a) If (x,y) and (x',y') are points of A in $Q_{i_1 \cdots i_s}$, then

$$|f(x',y') - f(x,y)| \leq 1/4^s . \tag{2}$$

(b) If (x,y) and (x',y') are points of A separated by some point $Q_{i_1 i_2 \cdots}$, and if $Q_{i_1 \cdots i_s}$ is the smallest square containing them both, then

$$r > \frac{1}{12}\frac{1}{3^{s+1}} . \tag{3}$$

Assume (a) and (b) are true. Given $\epsilon > 0$, choose s_0 and δ so that

$$36\left(\frac{3}{4}\right)^{s_0} < \epsilon, \qquad \delta < \frac{1}{12}\frac{1}{3^{s_0+1}} . \tag{4}$$

Now let (x,y) and (x',y') be any two points of A distant $r < \delta$. If no point $Q_{i_1 i_2 \cdots}$ separates them, then $f(x',y') = f(x,y)$, and (1) holds. Otherwise, let $Q_{i_1 \cdots i_s}$ be the smallest square containing them both. (b) gives

$$\frac{1}{12}\frac{1}{3^{s+1}} < r < \delta < \frac{1}{12}\frac{1}{3^{s_0+1}}, \quad \therefore s > s_0 .$$

Hence

$$\frac{|f(x',y') - f(x,y)|}{r} \leq \frac{12 \cdot 3^{s+1}}{4^s} = 36\left(\frac{3}{4}\right)^s < \epsilon .$$

It remains to prove (a) and (b). (a) is obvious from the definition of f. To prove (b), we consider three cases: neither point is in any square $Q_{i_1 \cdots i_s i_{s+1}}$; each point is in such a square; one point is in such a square, and the other is not. In the first two cases we see that $r > 1/(2 \cdot 3^{s+1})$; in the third case, (3) holds.

4. **Generalization to higher dimensions.** We shall indicate how the corresponding example is constructed for $n = 3$, $m = 2$; the generalization to higher dimensions is obvious.

Let Q be a cube of side 1. Let $Q_0, \cdots, Q_7$ be cubes of side 2/5 arranged in Q so that Q_i is adjacent to Q_{i-1}. Let q and q' be the centers of faces of Q which are adjacent and adjacent respectively to Q_0 and Q_7. Define q_i and q_i' $(i = 0, \cdots, 7)$ as before, and similarly for $A_0, \cdots, A_8$. The process is continued indefinitely, as before. A is again an arc. We set

$$\text{on } A_{i_1 \cdots i_s}, \quad f = \frac{i_1}{8} + \cdots + \frac{i_s}{8^s},$$

$$\text{at } Q_{i_1 i_2 \cdots}, \quad f = \frac{i_1}{8} + \frac{i_2}{8^2} + \cdots .$$

Set $f_{000} = f$, $f_{\alpha\beta\gamma} = 0$ $(\alpha + \beta + \gamma = 1 \text{ or } 2)$ on A. To prove that $f = f_{000}$ is of class C^2 in terms of these functions we need merely show that

$$\Delta = | f(x', y', z') - f(x, y, z) | < r^2\epsilon \text{ on } A \qquad (r < \delta) . \tag{5}$$

Note that f varies by at most $1/8^s$ in any $Q_{i_1 \cdots i_s}$. Now if the smallest cube containing (x,y,z) and (x',y',z') is $Q_{i_1 \cdots i_s}$, r is of the order $(2/5)^s$, while $\Delta \leqq 1/8^s$; hence Δ/r^2 is of the order $(25/32)^s$, which is $< \epsilon$ for r small enough and hence s large enough.

Note that *all partial derivatives of* f (of order $\leqq n - 1 = 2$) *vanish on* A.

HARVARD UNIVERSITY.

ONE-PARAMETER GROUPS OF TRANSFORMATIONS IN ABSTRACT VECTOR SPACES

By D. S. Nathan

1. Introduction. Consider an abstract function in the sense of an operation on real numbers which yields for each number an element of a linear, normed space. The derivative of such a function has been defined by Fréchet.[1] The same formal rules are valid as for the derivative of an ordinary function. Once the concept of a derivative is at hand, one can generalize from ordinary to abstract differential equations. The theory of Riemann integration of functions of a real variable whose values belong to a complete, linear, normed space[2] $\mathfrak{B}$ has been developed by Graves[3] and by Kerner.[4] This theory is analogous, on the whole, to the classical theory.

The notion of integral leads to existence theorems for abstract differential equations. By applying an existence theorem of this type due to Kerner, we find that a transformation in $\mathfrak{B}$ satisfying a Lipschitz condition generates, in the sense of integration of an abstract differential equation, a one-parameter group of transformations in $\mathfrak{B}$ or, at least, the group-germ. The group generated by a given bounded linear transformation is obtained as an exponential series. Conversely, the transformation which generates the subset in a certain neighborhood of the identity transformation of a given one-parameter group of bounded linear transformations with an additive law of composition is obtained as a logarithmic series. That this converse theorem can have content is made clear through the consideration of a certain class of one-parameter unitary groups in Hilbert space. In this connection use is made of results due to Stone.[5] While Stone's results are more powerful than ours in that the generating transformation need not be bounded, they are restricted to the case of unitary groups. Furthermore, even in this case it is of interest to fix the precise scope of the representation by power series.

Received April 17, 1935; presented to the American Mathematical Society, March 30, 1934. The author is a National Research Fellow.

[1] M. Fréchet, *La notion de différentielle dans l'analyse générale*, Annales Scientifiques de l'École Normale Supérieure, vol. 42 (1925), p. 312.

[2] S. Banach, *Théorie des Opérations Linéaires*, Warsaw, 1932, p. 53.

[3] L. M. Graves, *Riemann integration and Taylor's theorem in general analysis*, Transactions of the American Mathematical Society, vol. 29 (1927), pp. 163–177.

[4] M. Kerner, *Gewöhnliche Differentialgleichungen der allgemeinen Analysis*, Prace Mathematyczno-Fizyczne, vol. 40 (1932), pp. 47–67.

[5] M. H. Stone, *On one-parameter unitary groups in Hilbert space*, Annals of Mathematics, vol. 33 (1932), pp. 643–648.

2. **One-parameter groups.** Using the method of successive approximations, Kerner[6] has proved the following

THEOREM 1. *Let the operation $F(f, t)$ order an element of $\mathfrak{B}$ to each element f of a neighborhood of f_0 in $\mathfrak{B}$ and to each real number t in a neighborhood of t_0. Let F be continuous with respect to f and t for certain neighborhoods of f_0 and t_0. Let there be neighborhoods of f_0 and t_0 in which a Lipschitz condition holds of the type*

$$\| F(f, t) - F(g, t) \| \leqq C \| f - g \| ,$$

where C is a constant independent of f, g and t. Then there is a neighborhood of t_0 in which the equation

$$\frac{df}{dt} = F(f, t)$$

possesses a unique solution $f = \varphi(t; t_0, f_0)$, continuous in all arguments, such that $\varphi(t_0; t_0, f_0) = f_0$.[7]

Let T be a transformation with domain $\mathfrak{D}_0$ and range $\mathfrak{R}$[8] (where $\mathfrak{D}_0$ is an arbitrary open subset, and $\mathfrak{R}$ any subset, of $\mathfrak{B}$) such that for every pair of elements f and g in $\mathfrak{D}_0$ the inequality $\| Tf - Tg \| \leqq C \| f - g \|$ holds, C being a constant independent of f and g. Then T is continuous throughout $\mathfrak{D}_0$, since at any element f of $\mathfrak{D}_0$ there corresponds to every $\epsilon > 0$ a $\delta = \epsilon/C$ such that $\| Tf - Tg \| \leqq \epsilon$ when $\| f - g \| \leqq \delta$. Thus T meets the requirements for F in Theorem 1 throughout $\mathfrak{D}_0$. Hence, for $| t - t_0 |$ sufficiently small, the equation

$$\frac{df}{dt} = Tf \tag{1}$$

possesses a unique continuous solution $f = \varphi(t; t_0, f_0)$ satisfying the initial condition $\varphi(t_0; t_0, f_0) = f_0$, where f_0 is an arbitrary element of $\mathfrak{D}_0$. The solution can always be written in the form

$$f = \varphi(t - t_0; f_0) , \qquad f_0 = \varphi(0; f_0) , \tag{2}$$

as can be readily seen.

For every value of the parameter t in a certain neighborhood of t_0, (2) defines a transformation of f_0 into a point f of $\mathfrak{D}_0$. We have, then, a one-parameter family of transformations A_{t-t_0} of $\mathfrak{D}_0$ into a subset of $\mathfrak{D}_0$ defined by

$$f = A_{t-t_0} f_0 \equiv \varphi(t - t_0; f_0) ,$$

with the continuity property $A_{t-t_0} \to A_{t'-t_0}$ as $t \to t'$.[9]

[6] Loc. cit., pp. 60, 64. The proof is analogous to that usually encountered in the case of ordinary differential equations, as, for example, in L. Bieberbach, *Differentialgleichungen*, 1930, Pt. I, Ch. II.

[7] Instead of the Lipschitz condition Kerner has the hypothesis that F possesses a continuous differential in certain neighborhoods of f_0 and t_0, but from this he deduces the Lipschitz condition, and it is the latter that enters into the proof of the theorem.

[8] That is, let T be a correspondence between $\mathfrak{D}_0$ and $\mathfrak{R}$ whereby to each element of $\mathfrak{D}_0$ is ordered one element of $\mathfrak{R}$ and each element of $\mathfrak{R}$ is ordered to at least one element of $\mathfrak{D}_0$.

[9] Convergence will be understood always in the strong sense (Banach, loc. cit., p. 16).

The family A_{t-t_0} forms a group-germ. For let f_0 be defined by

$$f_0 = \varphi(t_1 - t_0; g) ,$$

where g is an arbitrary element of $\mathfrak{D}_0$ and $| t_1 - t_0 |$ is sufficiently small. Then $\varphi(t - t_0; g)$ is the solution of (1) passing through the point (t_1, f_0) in the product space formed by composition of $\mathfrak{B}$ and the real axis. Since, however, the unique solution passing through (t_1, f_0) is $\varphi(t - t_1; f_0)$, there holds identically for values of t and t_1 sufficiently near t_0

$$\varphi(t - t_1; \varphi(t_1 - t_0; g)) = \varphi(t - t_0; g) ,$$

that is,

$$A_{t-t_1}A_{t_1-t_0}g = A_{t-t_0}g .$$

It is evident from the initial condition that $A_0 = I$ (the identity transformation), and from the additive law of composition just obtained that A_{t-t_0} has an inverse A_{t_0-t}; in symbols, $A^{-1}_{t-t_0} = A_{t_0-t}$. The existence of an inverse tells us that the range of A_{t-t_0} is $\mathfrak{D}_0$. Multiplication of the transformations in the family is associative.

If we assume further that the domain of T is $\mathfrak{B}$ and that T satisfies for every f in $\mathfrak{B}$ the inequality $\| Tf \| \leqq M$, where M is a constant independent of f, the solution (2) can be extended over the infinite t-interval, so that A_{t-t_0} is defined for $-\infty < t < \infty$ and an actual group is generated.

Finally, there is no loss of generality if we take $t_0 = 0$.

We have, then, the following

THEOREM 2. *Let T be a transformation with domain $\mathfrak{D}_0$[10] satisfying for every f and g in $\mathfrak{D}_0$ the inequality $\| Tf - Tg \| \leqq C \| f - g \|$, where C is a constant independent of f and g. Then T generates the germ A_t of a one-parameter group of transformations of $\mathfrak{D}_0$ into $\mathfrak{D}_0$, with the law of composition $A_{t_1} A_{t_2} = A_{t_1+t_2}$ and the continuity property $A_t \to A_{t'}$ as $t \to t'$. If, further, the domain of T is $\mathfrak{B}$ and if, for every f in $\mathfrak{B}$, the inequality $\| Tf \| \leqq M$ holds, where M is a constant independent of f, then T generates the one-parameter group of transformations A_t, $-\infty < t < \infty$, with domain and range $\mathfrak{B}$.*

3. **One-parameter linear groups.** Let T be a bounded linear transformation with domain $\mathfrak{B}$, and let $| T |$ be the bound of T.[11] In view of the inequality $\| Tf - Tg \| = \| T(f - g) \| \leqq | T | \| f - g \|$, Theorem 2 is applicable, telling us that T generates the germ of a one-parameter group of transformations taking $\mathfrak{B}$ into $\mathfrak{B}$. We shall now obtain an explicit expression for A_t in terms

[10] The range of a transformation, when not expressly stated, is understood to be a subset of $\mathfrak{B}$.

[11] T is said to be linear if its domain includes $a_1f_1 + \cdots + a_kf_k$ ($a_1, \cdots, a_k$ arbitrary complex numbers) whenever it includes $f_1, \cdots, f_k$, and if then $T(a_1f_1 + \cdots + a_kf_k) = a_1Tf_1 + \cdots + a_kTf_k$. T is said to be bounded if there exists a constant C independent of f such that for every f in the domain of $\mathfrak{B}$ the inequality $\| Tf \| \leqq C \| f \|$ is valid. The smallest such C is called the bound of T and will be denoted by $| T |$.

of T. To this end, let us set down the sequence of approximations to the solution of the equation

$$\text{(3)} \qquad \frac{df}{dt} = Tf, \qquad f(0) = f_0 .$$

The sequence is as follows:

$$\varphi_0(t) = f_0 ,$$

$$\varphi_n(t) = f_0 + \int_0^t T\varphi_{n-1}(t)dt = \left(\sum_{\nu=0}^{n} \frac{t^\nu}{\nu!} T^\nu\right) f_0 \qquad (n = 1, 2, 3, \cdots),$$

where $T^0 = I, T^1 = T, T^n = TT^{n-1}$. The sequence

$$\{F_n(t; T)\} = \left\{\sum_{\nu=0}^{n} \frac{t^\nu}{\nu!} T^\nu\right\}$$

converges in $\mathfrak{B}$ for all t and uniformly in any bounded t-interval. For taking $m < n$, we have for every f in $\mathfrak{B}$

$$\| [F_n(t; T) - F_m(t; T)] f \| \leqq \sum_{\nu=m+1}^{n} \left\| \frac{t^\nu}{\nu!} T^\nu f \right\| \leqq \sum_{\nu=m+1}^{n} \frac{| T |^\nu | t^\nu |}{\nu!} \| f \|,$$

and since there corresponds to every $\epsilon > 0$ an $M = M(\epsilon)$ such that $\sum_{\nu=m+1}^{n} \frac{| T |^\nu | t^\nu |}{\nu!} < \epsilon (m > M)$, we infer the inequality $\| [F_n(t; T) - F_m(t; T)] f \| < \epsilon \| f \| (m > M)$. Since $\mathfrak{B}$ is complete, $\{F_n(t; T)\}$ converges to a transformation in $\mathfrak{B}$ which we denote by the symbol

$$e^{tT} \equiv \lim_{n\to\infty} \sum_{\nu=0}^{n} \frac{t^\nu}{\nu!} T^\nu \qquad (-\infty < t < \infty).$$

Then the unique solution of (3) is

$$f = \lim_{n\to\infty} \varphi_n(t) = e^{tT} f_0 \qquad (-\infty < t < \infty).$$

T, then, generates the one-parameter group $A_t \equiv e^{tT}, -\infty < t < \infty$.

The transformations constituting this group are linear. For hold t fixed and let $f_1, \cdots, f_k$ be arbitrary elements of $\mathfrak{B}$. We have

$F_n(t; T)(a_1 f_1 + \cdots + a_k f_k) = a_1 F_n(t; T) f_1 + \cdots + a_k F_n(t; T) f_k \to a_1 A_t f_1 + \cdots + a_k A_t f_k$ as $n \to \infty$; hence $A_t(a_1 f_1 + \cdots + a_k f_k) = a_1 A_t f_1 + \cdots + a_k A_t f_k$.

The transformations A_t are bounded. For hold t fixed and let f be any element in $\mathfrak{B}$. We have

$$\| A_t f \| \leqq \| A_t f - F_n(t; T) f \| + \| F_n(t; T) f \| .$$

To every $\epsilon > 0$ there corresponds an $N = N(\epsilon)$ such that $\|A_t f - F_n(t; T)f\| \leqq \epsilon \|f\|$ $(n > N)$; furthermore $\|F_n(t; T)f\| \leqq \sum_{\nu=0}^{n} (|T|^\nu/\nu!)\,|t^\nu|\,\|f\|$; hence

$$\|A_t f\| \leqq \left(\epsilon + \sum_{\nu=0}^{n} \frac{|T|^\nu |t^\nu|}{\nu!}\right) \|f\|.$$

We accordingly have

Theorem 3. *Let T be a bounded linear transformation with domain $\mathfrak{B}$. Then T generates the one-parameter group of bounded linear transformations of $\mathfrak{B}$ into $\mathfrak{B}$*

$$A_t \equiv e^{tT} \equiv \lim_{n \to \infty} \sum_{\nu=0}^{n} \frac{t^\nu}{\nu!} T^\nu \qquad (-\infty < t < \infty)$$

with the law of composition $A_{t_1}A_{t_2} = A_{t_1+t_2}$ and the continuity property $A_t \to A_{t'}$ as $t \to t'$.

The investigation of the converse problem involves a study of the logarithmic series

$$\sum_{\nu=1}^{\infty} \frac{(-1)^{\nu-1}}{\nu} (A - I)^\nu, \tag{4}$$

where A is a linear transformation with domain $\mathfrak{B}$. Set

$$G_n(A) = \sum_{\nu=1}^{n} \frac{(-1)^{\nu-1}}{\nu} (A - I)^\nu,$$

and let A satisfy for every f in $\mathfrak{B}$ the inequality $\|(A - I)f\| \leqq \theta \|f\|$, where θ is independent of f. Taking $m < n$, we have $\|[G_n(A) - G_m(A)]f\| \leqq \sum_{\nu=m+1}^{n} (\theta^\nu/\nu)\|f\|$. Corresponding to every $\epsilon > 0$ there exists an $M = M(\epsilon)$ such that $\sum_{\nu=m+1}^{n} \theta^\nu/\nu < \epsilon$ $(m > M, 0 \leqq \theta < 1)$. Hence

$$\|[G_n(A) - G_m(A)]f\| < \epsilon \|f\|$$
$$(m > M, \|(A - I)f\| \leqq \theta \|f\|, 0 \leqq \theta < 1).$$

Then if the inequality

$$\|(A - I)f\| \leqq \theta \|f\| \qquad (0 \leqq \theta < 1, \theta \text{ independent of } f) \tag{5}$$

holds for every f in $\mathfrak{B}$, the sequence $\{G_n(A)\}$ converges to a transformation in $\mathfrak{B}$ which we designate by

$$lg\, A \equiv \sum_{\nu=1}^{\infty} \frac{(-1)^{\nu-1}}{\nu} (A - I)^\nu.$$

The proof that lgA is linear and bounded is like the proof of these properties for e^T. We thus have

Lemma 1. *Let A be a linear transformation with domain $\mathfrak{B}$ such that $\|(A - I)f\| \leqq \theta \|f\|$ $(0 \leqq \theta < 1, \theta$ independent of $f)$ for every f in $\mathfrak{B}$. Then*

$$lg\ A = \sum_{\nu=1}^{\infty} \frac{(-1)^{\nu-1}}{\nu} (A - I)^{\nu}$$

is a bounded linear transformation with domain $\mathfrak{B}$.

That the condition (5) for the convergence of (4) throughout $\mathfrak{B}$ is the best possible is seen from the following example. Let $\mathfrak{B}$ be specialized to abstract Hilbert space $\mathfrak{H}$, and let $\{\varphi_n\}$ be a complete orthonormal set in $\mathfrak{H}$.[12] Define A as follows:

$$(A - I)\varphi_n = \theta_n \varphi_n, \qquad -1 < \theta_n < 0, \qquad n = 1, 2, 3, \cdots,$$

where A is taken as linear. Then A is bounded and has $\mathfrak{H}$ as its domain. If $f = \sum_{i=1}^{\infty} a_i \varphi_i$ is an arbitrary element of $\mathfrak{H}$, we have $\| f \|^2 = \sum_{i=1}^{\infty} | a_i |^2$, $\| (A - I)f \|^2 = \sum_{i=1}^{\infty} | a_i |^2 \theta_i^2$, therefore $\| (A - I)f \| < \| f \|$. Choose an element $f_1 = \sum_{i=1}^{\infty} b_i \varphi_i$, where at most a finite number of the b_i vanish. Using the relation $(A - I)^m \varphi_n = \theta_n^m \varphi_n$, we get, upon expanding $\sum_{\nu=1}^{\infty} [(-1)^{\nu-1} (A - I)^{\nu}/\nu] f_1$ and collecting the coefficients of the φ_i,

$$\left[\sum_{\nu=1}^{\infty} \frac{(-1)^{\nu-1}}{\nu} (A - I)^{\nu} \right] f_1 = \sum_{i=1}^{\infty} b_i\, lg\, (1 + \theta_i)\, \varphi_i .$$

By choosing the θ_i sufficiently near -1, the expressions $lg(1 + \theta_i)$ will become so large numerically that $\sum_{i=1}^{\infty} | b_i\, lg(1 + \theta_i) |^2$ will diverge; in other words, $\sum_{\nu=1}^{\infty} (-1)^{\nu-1} (A - I)^{\nu}/\nu$ will not be defined at f_1.

We shall require, further, certain properties of bounded linear transformations A and B with domain $\mathfrak{B}$. As the corresponding properties have been established for finite matrices by von Neumann,[13] we shall, for the most part, confine ourselves to indicating the slight modifications of his proofs which are entailed.

LEMMA 2. $\| (e^A - I)f \| \leqq (e^{|A|} - 1) \| f \|$.

Proof.

$$\| (e^A - I)f \| \leqq \sum_{\nu=1}^{\infty} \frac{1}{\nu !} \| A^{\nu} f \| \leqq \sum_{\nu=1}^{\infty} \frac{| A |^{\nu}}{\nu !} \| f \| .$$

LEMMA 3. *If for every f in $\mathfrak{B}$ the inequality*

$$\| (A - I)f \| < \omega \| f \|$$

($0 < \omega < 1$ *and ω independent of f) is satisfied, then*

$$\| (lg\ A)f \| < \left(lg \frac{1}{1 - \omega} \right) \| f \| .$$

[12] M. H. Stone, *Linear Transformations in Hilbert Space*, 1932, pp. 3, 7.

[13] J. von Neumann, *Über die analytischen Eigenschaften von Gruppen linearer Transformationen und ihrer Darstellungen*, Mathematische Zeitschrift, vol. 30 (1929), pp. 1–42.

Proof.

$$\| (lg\, A)f \| \leqq \sum_{\nu=1}^{\infty} \frac{1}{\nu} \| (A - I)^{\nu} f \| < \sum_{\nu=1}^{\infty} \frac{\omega^{\nu}}{\nu} \| f \| .$$

LEMMA 4.

(a) *If for every f in $\mathfrak{B}$ the inequality $\| (A - I)f \| \leqq \theta \| f \|$ $(0 \leqq \theta < 1$ and θ independent of $f)$ is satisfied, then $e^{lg\, A} = A$.*

(b) *If $| A | < lg\, 2$, then $lg\, e^{A} = A$.*[14]

The validity of (a) and (b) will be assured if it can be shown that corresponding to every $\epsilon > 0$ there exists a $t = t(\epsilon)$ such that

$$\left\| \sum_{\mu=0}^{r} \frac{1}{\mu!} \left[\sum_{\nu=1}^{s} \frac{(-1)^{\nu-1}}{\nu} (A - I)^{\nu} \right]^{\mu} f - Af \right\| < \epsilon ,$$

$$\left\| \sum_{\mu=1}^{r} \frac{(-1)^{\mu-1}}{\mu} \left(\sum_{\nu=1}^{s} \frac{A^{\nu}}{\nu!} \right)^{\mu} f - Af \right\| < \epsilon \qquad (s \geqq t, r \geqq t) .$$

The existence of such a t is proved in a manner analogous to von Neumann's proof of the corresponding result for finite matrices.[15]

LEMMA 5. *If A and B are permutable, then $e^{A}e^{B} = e^{A+B}$.*[16]

The proof is substantially the same as von Neumann's proof of the property for permutable finite matrices.[17]

LEMMA 6. *If A and B are permutable, and if for every f in $\mathfrak{B}$ the inequalities*

$$\| (A - I)f \| \leqq \theta \| f \| , \qquad \| (B - I)f \| \leqq \theta \| f \| ,$$

$\| (AB - I)f \| \leqq \theta \| f \|$ $(0 \leqq \theta < 1$, θ independent of $f)$ are satisfied, then

$$lg\, AB = lg\, A + lg\, B . \tag{6}$$

The proof, again, is analogous to von Neumann's proof of the property for permutable finite matrices.[18] In the first place, it can be demonstrated that if (6) holds when both $\| (A - I)f \| < \delta \| f \|$ and $\| (B - I)f \| < \delta \| f \|$ for every f in $\mathfrak{B}$, where δ is some constant independent of f, then (6) holds under the conditions stated in the lemma. But (6) actually is valid if for every f in $\mathfrak{B}$ the transformations A and B satisfy the inequalities

$$\| (A - I)f \| < \omega \| f \| , \quad \| (B - I)f \| < \omega \| f \| \quad (0 < \omega < 1 - \sqrt{1/2}), \tag{7}$$

with ω independent of f.

[14] That $lg\, e^{A}$ is defined for $| A | < lg\, 2$ follows from the inequality (arising from Lemma 2) $\|(e^{A} - I)f \| \leqq (e^{|A|} - 1) \| f \|$, where $0 \leqq e^{|A|} - 1 < 1$.

[15] Loc. cit., pp. 9–10.

[16] This result yields an alternative proof that $e^{t_1 T} e^{t_2 T} = e^{(t_1 + t_2)T}$.

[17] Loc. cit., p. 10.

[18] Loc. cit., p. 11.

For by Lemma 3 we have, when (7) is satisfied,

$$\|(lg\ A)f\| < \left(lg \frac{1}{1-\omega}\right)\|f\|, \qquad \|(lg\ B)f\| < \left(lg \frac{1}{1-\omega}\right)\|f\|,$$

and consequently

$$\|(lg\ A + lg\ B)f\| < \left(2\, lg \frac{1}{1-\omega}\right)\|f\|,$$

where $2\, lg\,[1/(1-\omega)] < lg\ 2$. By Lemma 4(b), then, we have

$$lg\ e^{lg\,A + lg\,B} = lg\ A + lg\ B\,. \tag{8}$$

As limits of polynomials in A and B, $lg\ A$ and $lg\ B$ are permutable. By Lemmas 5 and 4(a) we get $e^{lg\,A+lg\,B} = e^{lg\,A}\,e^{lg\,B} = AB$, whence $lg\ e^{lg\,A+lg\,B} = lg\ AB$. Comparing this with (8) we get (6); in other words, (6) holds if (7) is satisfied. Hence the lemma is proved.

We are now in a position to prove the following converse of Theorem 3.

THEOREM 4. *Let A_t, $-\infty < t < \infty$, be a one-parameter group of bounded linear transformations with domain $\mathfrak{B}$, with the law of composition $A_{t_1}A_{t_2} = A_{t_1+t_2}$ and the continuity property $A_t \to A_{t'}$ as $t \to t'$. Let $\mathfrak{T}$ denote the t-interval such that for every f in $\mathfrak{B}$ and every pair of values σ and τ in $\mathfrak{T}$ the inequality $\|(A_{\sigma+\tau} - I)f\| \leqq \theta\|f\|$ $(0 \leqq \theta < 1$, θ independent of $f)$ is satisfied. Then the subset of A_t corresponding to the values of t in $\mathfrak{T}$ is generated by the bounded linear transformation $(lg\ A_t)/t$ with domain $\mathfrak{B}$, which is independent of t.*

Proof. We show first that $(lg\ A_t)/t$ is independent of t. By hypothesis and Lemma 6 we have

$$lg\ A_{t_1+t_2} = lg\ A_{t_1}A_{t_2} = lg\ A_{t_1} + lg\ A_{t_2}\,, \tag{9}$$

if the following inequalities hold for every f in $\mathfrak{B}$:

$$\|(A_{t_1} - I)f\| \leqq \theta\|f\|, \quad \|(A_{t_2} - I)f\| \leqq \theta\|f\|, \quad \|(A_{t_1+t_2} - I)f\| \leqq \theta\|f\|$$
$$(0 \leqq \theta < 1 \text{ and } \theta \text{ independent of } f)\,.$$

Then (9) is valid if t_1 and t_2 are both contained in $\mathfrak{T}$. Put $(lg\ A_t)f = g(t)$. Since $lg\ A_t$ is continuous in t, $g(t)$ is a continuous abstract function of t whose values belong to the subset $\mathfrak{R}'$ of $\mathfrak{B}$ which is the range of $lg\ A_t$. We get from (9)

$$g(t_1 + t_2) = g(t_1) + g(t_2)\,. \tag{10}$$

Using the precise form of argument by which Cauchy[19] proved that the solution of the equation

$$F(x + y) = F(x) + F(y)\,,$$

where $F(x)$ is a real continuous function of a real variable x, is $F(x) = mx$ (m an arbitrary real constant), we find that the solution of (10) is $g(t) = th$ (h an arbi-

[19] A. Cauchy, *Oeuvres Complètes*, 2nd series, vol. 3 (1897), pp. 99–100.

trary element of $\mathfrak{R}'$). Then $g(t)/t = h$ $(t \neq 0)$, i.e., $[(lg\, A_t)/t]f = h$ $(t \neq 0)$. Hence $(lg\, A_t)/t$ $(t \neq 0)$ is independent of t if t is in $\mathfrak{T}$. It is obvious that $(lg\, A_t)/t$ approaches a limit as $t \to 0$. Since $lg\, A_t$ is linear and bounded and has $\mathfrak{B}$ as its domain for the values of t for which it is defined (Lemma 1), $(lg\, A_t)/t$ is linear and bounded and has $\mathfrak{B}$ as its domain for the values of t in $\mathfrak{T}$. By Theorem 3 and Lemma 4(a), the transformation $(lg\, A_t)/t$ generates $e^{lg_{*}A_t} = A_t$ for the values of t in $\mathfrak{T}$. This completes the proof of the theorem.

To show that Theorem 4 is not vacuous, we shall exhibit a class of one-parameter groups for which the interval $\mathfrak{T}$ exists. Let H be a bounded self-adjoint[20] transformation in $\mathfrak{H}$. By Theorem 3, the transformation iH $(i = \sqrt{-1})$ generates the one-parameter group e^{itH}, $-\infty < t < \infty$, with domain and range $\mathfrak{H}$. Our symbol

$$e^{itH} \equiv \lim_{n\to\infty} \sum_{\nu=0}^{n} \frac{(itH)^{\nu}}{\nu\,!}$$

can readily be seen to be identical with the symbol from the operational calculus

$$e^{itH} \equiv \int_{-\infty}^{\infty} e^{it\lambda}\, dE(\lambda)\,, \tag{11}$$

where $E(\lambda)$ is the resolution of the identity corresponding to H.[21] The transformations (11) are unitary, as Stone[22] has proved. Putting $U_t = e^{itH}$, we have

$$\|\,(U_t - I)f\,\|^2 = \int_{-\infty}^{\infty} |\,e^{it\lambda} - 1\,|^2\, d\,\|\,E(\lambda)f\,\|^2 = \int_{-|H|}^{|H|} |\,e^{it\lambda} - 1\,|^2\, d\,\|\,E(\lambda)f\,\|^2\,.$$

On a bounded λ-interval $e^{it\lambda} \to 1$ uniformly as $t \to 0$; that is, corresponding to every $\epsilon > 0$ there exists a $\delta = \delta(\epsilon)$ such that $|\,e^{it\lambda} - 1\,| \leqq \epsilon$ when $|\,t\,| \leqq \delta$. Choosing $\epsilon = \theta$, where θ is any positive number < 1, we have

$$\|\,(U_t - I)\,f\,\|^2 \leqq \theta^2 \int_{-|H|}^{|H|} d\,\|\,E(\lambda)\,f\,\|^2 = \theta^2\,\|\,f\,\|^2 \qquad (\,|\,t\,| \leqq \delta)\,,$$

$$\|\,(U_{t_1+t_2} - I)\,f\,\| \leqq \theta\,\|\,f\,\| \qquad \left[\max\,(\,|\,t_1\,|\,,\,|\,t_2\,|\,) \leqq \frac{\delta}{2}\right].$$

Thus $\mathfrak{T}$ exists for the unitary group e^{itH}.

Harvard University, Institute for Advanced Study, and Princeton University.

[20] Stone, *Linear Transformations in Hilbert Space*, p. 50.

[21] For terminology see Stone, *Linear Transformations in Hilbert Space*, Chapters V and VI. The second symbol has the advantage that it is defined also for non-bounded self-adjoint H.

[22] Stone, Annals of Mathematics, vol. 33, p. 643, Theorem A.

ON THE NEIGHBORHOOD OF A GEODESIC IN RIEMANNIAN SPACE

By J. L. Synge

1. Simple proof of a fundamental theorem in a space of positive curvature. Myers[1] has recently shown that a complete Riemannian manifold V_N, all of whose Riemannian curvatures are greater than a positive constant K_0, is closed and has a diameter less than $\pi K_0^{-1/2}$. The fundamental lemma on which this conclusion is based is the following:

Theorem I. *If all Riemannian curvatures of V_N are greater than or equal to a positive constant K_0, no geodesic arc of length greater than $\pi K_0^{-1/2}$ is the shortest curve joining the end points.*

In Myers' paper, and throughout the present paper, the line-element

$$ds^2 = a_{ij}\, dx^i\, dx^j \tag{1.1}$$

is assumed to be positive definite.

Myers' proof of Theorem I is similar to that given by Schoenberg,[2] and depends on the theory of conjugate points in the N-dimensional sense. However, the theorem is an immediate consequence of a result established by me,[3] which involves only the simpler concept of 2-dimensional conjugate points. But indeed the introduction of the idea of conjugate points and proofs of their existence are quite unnecessary for the establishment of Theorem I, as will now be shown.

Let AB be a geodesic arc of length L. Applying any infinitesimal variation η^i which vanishes at the end points, the first variation is of course zero, and the second variation is[4]

$$\delta^2 L = \frac{1}{2}\int_0^L [\eta'^2 + (\bar{\mu}^2 - K)\eta^2]ds\,, \tag{1.2}$$

where η is the magnitude of η^i, $\eta' = d\eta/ds$, $\bar{\mu}$ is the magnitude of the absolute derivative of the unit vector μ^i co-directional with η^i, and K is the Riemannian curvature of V_N for the 2-element containing the tangent to the geodesic and η^i. We may choose the unit vector μ^i as we like along AB, and we can assign η as we like, provided that $\eta = 0$ at A and at B. Let μ^i be propagated parallelly, so that $\bar{\mu} = 0$. Then, since by hypothesis $K \geqq K_0$, we have

$$\delta^2 L = \frac{1}{2}\int_0^L (\eta'^2 - K\eta^2)\, ds \leqq \frac{1}{2}\int_0^L (\eta'^2 - K_0\eta^2)\, ds\,. \tag{1.3}$$

Received May 15, 1935.

[1] S. B. Myers, this Journal, vol. 1 (1935), p. 42.

[2] I. J. Schoenberg, Annals of Math., vol. 33 (1932), p. 493.

[3] J. L. Synge, Proc. London Math. Soc., vol. 25 (1925), p. 264, Theorem XVII.

[4] Ref. 3, p. 261, equation (9.17); the notation is slightly changed.

Let us take

$$\eta = \epsilon \sin(\pi s/L) , \tag{1.4}$$

where ϵ is a positive constant. Then

$$\delta^2 L \leqq \frac{1}{2}\epsilon^2 \int_0^L \left(\frac{\pi^2}{L^2}\cos^2\frac{\pi s}{L} - K_0 \sin^2\frac{\pi s}{L}\right) ds = \frac{1}{4}\epsilon^2 L \left(\frac{\pi^2}{L^2} - K_0\right). \tag{1.5}$$

Thus the second variation is negative if

$$L > \pi K_0^{-1/2} , \tag{1.6}$$

which proves Theorem I.

Incidentally it may be remarked that a result established by Myers[5]—that the curvature along a geodesic g_0 of a V_2 formed by parallel propagation of an orthogonal geodesic g along g_0 is equal to the corresponding Riemannian curvature of V_N—is a particular case of a more general result established by me.[6]

2. **Geodesic deviation and its representation in the normal vector space.** In studying the properties of Riemannian manifolds in the large, it is of the greatest importance to know the behavior of geodesics drawn from a point in adjacent directions. If the manifold is connected and unbounded, there exists a geodesic joining any two points, and hence the congruence of geodesics from any point O contains all the points of the manifold.

For the investigation of the behavior of adjacent geodesics, the fundamental differential equation is[7,8]

$$\frac{\delta^2 \eta^i}{\delta s^2} + R^i_{.jkl}\, \lambda^j \eta^k \lambda^l = 0 , \tag{2.1}$$

where η^i is the infinitesimal displacement vector, perpendicular to the two adjacent geodesics in question, $R^i_{.jkl}$ the curvature tensor, and λ^i the unit tangent vector to one of the geodesics g_0; $\delta/\delta s$ is the operation of absolute differentiation along g_0. This equation was given independently by Levi-Civita,[9] who also considered a more general correspondence between the two geodesics.

If we introduce along g_0 $N - 1$ unit vectors $\lambda^i_{(\alpha)}$, orthogonal to one another and to the tangent λ^i, and allow Greek suffixes the range $1, 2, \cdots, N - 1$, with summation over the same range when repeated, we may write

$$\eta^i = X_\beta \lambda^i_{(\beta)} . \tag{2.2}$$

Substitution in (2.1) gives

[5] Ref. 1, p. 44.

[6] Ref. 3, p. 262, equation (9.19).

[7] J. L. Synge, Phil. Trans. Roy. Soc., London, A, vol. 226 (1926), p. 102, equation (9.12).

[8] J. L. Synge, Annals of Math., vol. 35 (1934), p. 707, equation (2.9).

[9] T. Levi-Civita, Math. Annalen, vol. 97 (1926), p. 315.

$$X''_\beta \lambda^i_{(\beta)} + 2X'_\beta \frac{\delta \lambda^i_{(\beta)}}{\delta s} + X_\beta \frac{\delta^2 \lambda^i_{(\beta)}}{\delta s^2} + X_\beta R^i_{.jkl} \lambda^j \lambda^k_{(\beta)} \lambda^l = 0 , \tag{2.3}$$

where the accents denote ordinary differentiation with respect to s. Since

$$\lambda^i_{(\alpha)} \lambda_{(\beta) i} = \delta_{\alpha\beta} , \tag{2.4}$$

the Kronecker delta, multiplication of (2.3) by $\lambda_{(\alpha) i}$ gives

$$X''_\alpha + 2X'_\beta \lambda_{(\alpha) i} \frac{\delta \lambda^i_{(\beta)}}{\delta s} + X_\beta \left(\lambda_{(\alpha) i} \frac{\delta^2 \lambda^i_{(\beta)}}{\delta s^2} + K_{\alpha\beta} \right) = 0 , \tag{2.5}$$

where

$$K_{\alpha\beta} = R_{ijkl} \lambda^i_{(\alpha)} \lambda^j \lambda^k_{(\beta)} \lambda^l = K_{\beta\alpha} ; \tag{2.6}$$

when $\alpha = \beta$, $K_{\alpha\beta}$ is the Riemannian curvature of V_N for the 2-element defined by $\lambda^i_{(\alpha)}$ and λ^i.

The mode of propagation of $\lambda^i_{(\alpha)}$ along g_0 is at our disposal, subject to the conditions of unit magnitude, mutual orthogonality and orthogonality to λ^i. On account of this last condition and the geodesic character of g_0, we have

$$\frac{\delta \lambda^i_{(\alpha)}}{\delta s} \lambda_i = 0 , \tag{2.7}$$

and consequently $\Omega_{\alpha\beta}$ exist, such that

$$\frac{\delta \lambda^i_{(\alpha)}}{\delta s} = \Omega_{\alpha\beta} \lambda^i_{(\beta)} . \tag{2.8}$$

From (2.4) it follows that

$$\Omega_{\beta\alpha} = - \Omega_{\alpha\beta} . \tag{2.9}$$

We have

$$\begin{aligned} \lambda_{(\alpha) i} \frac{\delta \lambda^i_{(\beta)}}{\delta s} &= \Omega_{\beta\gamma} \lambda^i_{(\gamma)} \lambda_{(\alpha) i} = \Omega_{\beta\alpha} , \\ \lambda_{(\alpha) i} \frac{\delta^2 \lambda^i_{(\beta)}}{\delta s^2} &= \lambda_{(\alpha) i} (\Omega'_{\beta\gamma} \lambda^i_{(\gamma)} + \Omega_{\beta\gamma} \Omega_{\gamma\delta} \lambda^i_{(\delta)}) \\ &= \Omega'_{\beta\alpha} + \Omega_{\beta\gamma} \Omega_{\gamma\alpha} , \end{aligned} \tag{2.10}$$

and so (2.5) gives

$$X''_\alpha + 2X'_\beta \Omega_{\beta\alpha} + X_\beta (\Omega'_{\beta\alpha} + \Omega_{\beta\gamma} \Omega_{\gamma\alpha} + K_{\alpha\beta}) = 0 . \tag{2.11}$$

We may choose $\Omega_{\alpha\beta}$ as arbitrary functions of s to suit our convenience, but it seems best to let $\lambda^i_{(\alpha)}$ undergo parallel propagation along $g_0 (\Omega_{\alpha\beta} = 0)$. Then (2.5) becomes[10]

$$X''_\alpha + K_{\alpha\beta} X_\beta = 0 . \tag{2.12}$$

[10] Ref. 7, p. 105, equation (9.52); ref. 9, p. 319. This is the same as Myers' equation (3.2), ref. 1.

The quantities X_α may be regarded as the coördinates in a flat space of $N-1$ dimensions, in which $\lambda^i_{(\alpha)}$ are axes of coördinates. We shall call this the *normal vector space*. In studying the behavior of geodesics neighboring the geodesic g_0, we may employ this normal vector space as a representative space. Each neighboring geodesic is represented by a moving point in this space, and g_0 is represented permanently by the origin. The $\frac{1}{2}N(N-1)$ quantities $\Omega_{\alpha\beta}$ are invariants with respect to transformations of the coördinates x^i in V_N: they are the components of a skew-symmetric cartesian tensor with respect to orthogonal transformations in the normal vector space, being in fact the components of *angular velocity* of the axes.

We shall call *principal curvature directions* those directions normal to g_0 for which the Riemannian curvature corresponding to g_0 and each of the directions in question has a stationary value, and we shall call the corresponding 2-elements *principal 2-elements.* If $\lambda^i_{(\alpha)}$ are arbitrarily assigned, the Riemannian curvature corresponding to g_0 and the direction defined by X_α is

$$K = \frac{K_{\alpha\beta}\, X_\alpha X_\beta}{X_\gamma X_\gamma}\,. \tag{2.13}$$

The $N-1$ principal curvature directions, which are, of course, mutually orthogonal, satisfy

$$\theta X_\alpha = K_{\alpha\beta} X_\beta\,, \tag{2.14}$$

where θ (the principal curvature) is a root of the characteristic determinantal equation

$$|\,\theta\delta_{\alpha\beta} - K_{\alpha\beta}\,| = 0\,. \tag{2.15}$$

If we choose the vectors $\lambda^i_{(\alpha)}$ in the principal curvature directions at a point on g_0, then at that point the matrix $K_{\alpha\beta}$ becomes diagonal, and at that point the variables are formally separated in (2.12). But in general the vectors $\lambda^i_{(\alpha)}$ will not remain principal as they undergo parallel propagation along g_0, and if we wish to retain a diagonal form for $K_{\alpha\beta}$, we are compelled to revert to (2.11) with appropriate values for $\Omega_{\alpha\beta}$. In general, it seems best to retain parallel propagation of $\lambda^i_{(\alpha)}$ and not attempt to obtain a diagonal form for $K_{\alpha\beta}$.

It will be observed that (2.12) are of dynamical form, s being interpreted as the "time". They are, in fact, the equations of motion of a particle of unit mass with kinetic and potential energies

$$T = \tfrac{1}{2}\, X'_\alpha X'_\alpha\,, \qquad V = \tfrac{1}{2}\, K_{\alpha\beta} X_\alpha X_\beta\,; \tag{2.16}$$

V in general involves the time s explicitly in $K_{\alpha\beta}$.

3. **Deviation from a steady geodesic.** In seeking a useful definition of steady motion,[11] I was led to consider a geodesic enjoying the property that $K_{\alpha\beta}$

[11] Ref. 7, p. 77.

are constant along it. This definition may also be stated in the more obviously invariant equivalent form: *A geodesic is steady if the principal curvature directions are propagated parallelly along it and the principal curvatures are constant.*

Since in such cases the potential energy V in (2.16) does not involve the time s explicitly, the theory of the deviation of geodesics adjacent to a steady geodesic is simply the well-known theory of the motion of a conservative dynamical system near a position of equilibrium. By choosing $\lambda^i_{(\alpha)}$ in the principal directions of curvature, X_α become normal coördinates, and (2.12) become

$$(3.1) \quad X_1'' + K_1X_1 = 0\,, \quad X_2'' + K_2X_2 = 0\,, \quad \cdots\,, \quad X_{N-1}'' + K_{N-1}X_{N-1} = 0\,,$$

where $K_1, K_2, \cdots, K_{N-1}$ are the principal curvatures.

The fundamental differential equation (2.1) is obtained from geometrical considerations by neglect of infinitesimals of order higher than the first. In stating deductions from this equation, we may choose between exact analytical statements concerning its solutions and approximate statements (properly interpreted) concerning the geometrical behavior of the geodesics. The latter course appears more suggestive geometrically, and we shall follow it. Thus two adjacent geodesics will be said to "intersect to the first order" at a point if the distance between the geodesics there is an infinitesimal of order higher than the first, the fundamental infinitesimal being the angle between the geodesics or their distance apart at some general assigned place.

The following results are immediate.

THEOREM II. *Let a geodesic g cut a steady geodesic g_0 at a small angle θ at $s = 0$. If the tangent to g at $s = 0$ lies in a principal 2-element with positive curvature K, then g lies permanently in the neighborhood of g_0 and intersects it to the first order at the points $s = n\pi K^{-1/2}(n = 0, \pm 1, \cdots)$. If the tangent to g at $s = 0$ has a non-zero component in any principal 2-element for which the principal curvature is zero or negative, g will not lie permanently in the neighborhood of g_0.*

THEOREM III. *If on a steady geodesic g_0 the principal curvatures are all positive (or, equivalently, if $K_{\alpha\beta}X_\alpha X_\beta$ is positive definite), any geodesic cutting g_0 at $s = 0$ at a small angle θ lies permanently in the neighborhood of g_0, the normal distance from g_0 never exceeding*

$$(3.2) \qquad \theta K_0^{-1/2}\,,$$

where K_0 is the smallest of the principal curvatures. The representative point never passes outside the quadric

$$(3.3) \qquad K_{\alpha\beta}X_\alpha X_\beta = \theta^2$$

in the normal vector space.

To see this, we note that in the case of a steady geodesic (2.12) possess the first integral

$$(3.4) \qquad X_\alpha' X_\alpha' + K_{\alpha\beta}X_\alpha X_\beta = 2E\,,$$

a constant, and $2E = \theta^2$, since at $s = 0$ we have $X_\alpha = 0$, $X'_\alpha X'_\alpha = \theta^2$. Thus

$$K_{\alpha\beta}X_\alpha X_\beta = \theta^2 - X'_\alpha X'_\alpha \leqq \theta^2 . \tag{3.5}$$

The value (3.2) is the length of the greatest radius vector that can be drawn from the origin to the quadric (3.3).

If the principal curvatures are positive, and normal coördinates are used, so that (3.1) hold, the solutions for a geodesic g intersecting g_0 at $s = 0$ at a small angle θ are

$$X_1 = A_1 \sin sK_1^{1/2}, \quad X_2 = A_2 \sin sK_2^{1/2}, \cdots, X_{N-1} = A_{N-1} \sin sK_{N-1}^{1/2}, \tag{3.6}$$

where the A's are constants such that

$$A_1^2K_1 + A_2^2K_2 + \cdots + A_{N-1}^2K_{N-1} = \theta^2 . \tag{3.7}$$

In view of the sense in which the "intersection" of geodesics is understood, the points on g_0 *conjugate to* $s = 0$ are the points of intersection to the first order of g_0 with geodesics g drawn through $s = 0$, making small angles with g_0. The following result holds.

THEOREM IV. *On a steady geodesic the points conjugate to $s = 0$ are those points for which s has the real values contained in the set of quantities*

$$n\pi K_1^{-1/2}, \quad n\pi K_2^{-1/2}, \quad \cdots, \quad n\pi K_{N-1}^{-1/2}, \qquad (n = \pm 1, \pm 2, \cdots) . \tag{3.8}$$

The points are the points of intersection to the first order of g_0 with geodesics g emanating in principal 2-elements at $s = 0$. Geodesics g emanating in other directions cut g_0 in some of the aforesaid points or do not cut g_0 at all. We may state the following result:

THEOREM V. *If g_0 is a steady geodesic for which the principal curvatures are positive and the ratios of their square roots irrational, no geodesic g cutting g_0 at a small angle at $s = 0$ ever cuts g_0 again to the first order, unless the direction of g at the point of intersection lies in one of the principal 2-elements.*

To trace the behavior of g, we return to (3.6). It is evident that the representative point in the normal vector space never passes outside the cuboidal region

$$-A_1 \leqq X_1 \leqq A_1, \quad -A_2 \leqq X_2 \leqq A_2, \quad \cdots, \quad -A_{N-1} \leqq X_{N-1} \leqq A_{N-1}. \tag{3.9}$$

The curve described by the representative point makes contact periodically with each of the $(N-2)$-flats

$$X_1 = \pm A_1, \qquad X_2 = \pm A_2, \qquad \cdots, \qquad X_{N-1} = \pm A_{N-1}, \tag{3.10}$$

which bound this region, and in the case where the ratios of the square roots of the principal curvatures are irrational the curve fills the region, in the sense that it passes as close as we like to any assigned point in it.

Let us put

$$X^2 = X_\alpha X_\alpha , \qquad X \geqq 0, \tag{3.11}$$

so that X is the distance of a point from the origin in the normal vector space; in fact, $X = \eta$, the distance between the two geodesics in question. Then

$$XX' = X_\alpha X'_\alpha\,, \tag{3.12}$$
$$XX'' + X'^2 = X'_\alpha X'_\alpha + X_\alpha X''_\alpha = X'_\alpha X'_\alpha - K_{\alpha\beta} X_\alpha X_\beta\,,$$

or by (3.4),

$$XX'' + X'^2 = 2E - 2K_{\alpha\beta} X_\alpha X_\beta\,. \tag{3.13}$$

Thus at a point where $X' = 0$, X'' has the same sign as

$$E - K_{\alpha\beta} X_\alpha X_\beta\,.$$

Referring to Theorem III, this result follows:[12]

THEOREM VI. *If a geodesic g cuts at a small angle θ a steady geodesic g_0 for which the principal curvatures are positive, all the points in the normal vector space for which the distance of g from g_0 has a minimum value lie inside the ellipsoid*

$$K_{\alpha\beta} X_\alpha X_\beta = \tfrac{1}{2}\,\theta^2\,, \tag{3.14}$$

and all the points for which the distance of g from g_0 has a maximum value lie in the homoeoid bounded by the ellipsoids

$$K_{\alpha\beta} X_\alpha X_\beta = \tfrac{1}{2}\,\theta^2\,, \qquad K_{\alpha\beta} X_\alpha X_\beta = \theta^2\,. \tag{3.15}$$

The curve in the normal vector space weaves in and out across the ellipsoid (3.14).

The special assumption on which the results of this section are based, namely, that the geodesic g_0 is steady, is, of course, of a very particular nature. The results are, however, of interest, as indicating the minimum degree of complexity to be expected in the general case when the special assumption does not hold.

4. The existence of conjugate points. We now return to the general case in which the special assumptions concerning the curvatures along the geodesic g_0 are not made. The fundamental differential equation for deviation is (2.12), that is,

$$X''_\alpha + K_{\alpha\beta} X_\beta = 0\,, \tag{4.1}$$

in which $K_{\alpha\beta}$ $(= K_{\beta\alpha})$ are to be regarded as arbitrary functions of the parameter s.

Let us consider the totality of geodesics emanating from the point $s = 0$ on g_0, making with g_0 all small angles less than $\bar{\theta}$, say. These geodesics are represented in the normal vector space by a cloud of moving points, which at "time" $s = 0$ are all at the origin and moving out from it with all velocities less than $\bar{\theta}$. We wish to study the behavior of this cloud. The argument

[12] Cf. J. L. Synge, Trans. Amer. Math. Soc., vol. 34 (1932), p. 494, Theorem IV.

which follows may be stated in a purely analytical form, or in a dynamical form, or in a hydrodynamical form. The last will be adopted.

Throughout the cloud there is a velocity distribution X'_α, these components being functions of position X_α and time s. The circulation in any closed circuit C, defined as

$$\int_C X'_\alpha dX_\alpha\,, \tag{4.2}$$

remains constant as the circuit moves with the fluid. For the equations of motion (4.1) may be put in Hamiltonian form

$$\begin{gathered} X'_\alpha = \partial H/\partial P_\alpha\,, \qquad P'_\alpha = -\,\partial H/\partial X_\alpha\,, \\ H = \tfrac{1}{2}\,P_\alpha P_\alpha + \tfrac{1}{2}\,K_{\alpha\beta}X_\alpha X_\beta\,, \end{gathered} \tag{4.3}$$

and the circulation is

$$\int_C P_\alpha dX_\alpha\,, \tag{4.4}$$

the well-known relative integral invariant.[13] Since every closed circuit was collapsed at the origin at $s = 0$, it follows that *the circulation in every closed circuit is zero*, and hence

$$\phi = \int_O^M X'_\alpha dX_\alpha\,, \tag{4.5}$$

the integral being taken from the origin O to any point $M(X_\alpha)$, is a single-valued function of X_α and s; in fact, hydrodynamically speaking, the motion of the cloud is *irrotational*, and ϕ is the *velocity potential* such that

$$X'_\alpha = \partial\phi/\partial X_\alpha\,. \tag{4.6}$$

This irrotational property depends on the fact that the geodesics in question all cut g_0 at a point; other special initial conditions would also lead to it.

Substitution from (4.6) in (4.1) gives

$$\frac{\partial^2\phi}{\partial X_\beta\partial X_\alpha}\,\frac{\partial\phi}{\partial X_\beta} + \frac{\partial^2\phi}{\partial s\partial X_\alpha} + K_{\alpha\beta}X_\beta = 0\,, \tag{4.7}$$

or

$$\frac{\partial}{\partial X_\alpha}\left(\frac{\partial\phi}{\partial s} + \frac{1}{2}\,\frac{\partial\phi}{\partial X_\beta}\,\frac{\partial\phi}{\partial X_\beta} + \frac{1}{2}\,K_{\beta\gamma}X_\beta X_\gamma\right) = 0\,, \tag{4.8}$$

so that

$$\frac{\partial\phi}{\partial s} + \frac{1}{2}\,\frac{\partial\phi}{\partial X_\beta}\,\frac{\partial\phi}{\partial X_\beta} + \frac{1}{2}\,K_{\beta\gamma}X_\beta X_\gamma = F(s)\,. \tag{4.9}$$

[13] Cf. E. T. Whittaker, *Analytical Dynamics*, Cambridge, 1927, p. 272.

Putting $X_\alpha = 0$, so that, in consequence,

$$\partial\phi/\partial s = 0 , \qquad \partial\phi/\partial X_\alpha = 0 ,$$

we see that $F(s) = 0$, and we have in general

$$\frac{\partial\phi}{\partial s} + \frac{1}{2}\frac{\partial\phi}{\partial X_\beta}\frac{\partial\phi}{\partial X_\beta} + \frac{1}{2}K_{\beta\gamma}X_\beta X_\gamma = 0 . \tag{4.10}$$

This is in fact the Hamilton-Jacobi equation,[14] in which the function ϕ has a simple kinematical meaning in our hydrodynamical theory. To sum up:

THEOREM VII. *For the congruence of geodesics intersecting a geodesic g_0 at a point at small angles, the motion of the corresponding cloud in the normal vector space is irrotational, and the velocity potential ϕ, defined as the instantaneous flow along an arbitrary curve drawn from the origin to a point X_α, satisfies the Hamilton-Jacobi equation* (4.10).

The velocity potential ϕ is capable of a still simpler interpretation. From the nature of the initial conditions, it follows that if $X_\alpha = f_\alpha(s)$ is a solution of (4.1) corresponding to some geodesic of the congruence, then $X_\alpha = Cf_\alpha(s)$ is also a solution belonging to the congruence, C being a constant. Giving any particular value to s, and letting C range, we get a straight line of points through the origin of the normal vector space, the corresponding velocities being parallel to one another and their magnitudes being proportional to X, the distance from the origin. Thus integrating along the straight line from the origin to any point P, we have

$$\phi_P = \int_O^P X'_\alpha dX_\alpha = \int_O^P (X'_\alpha)_P \frac{X}{X_P}\left(\frac{X_\alpha}{X}\right)_P dX = \frac{1}{2}(X'_\alpha X_\alpha)_P . \tag{4.11}$$

Now the *outward radial velocity* in the cloud is

$$v_r = X'_\alpha X_\alpha / X , \tag{4.12}$$

and hence *the velocity potential ϕ is connected with the outward radial velocity by the equation*

$$\phi = \tfrac{1}{2}Xv_r , \tag{4.13}$$

X being the distance from the origin.

Denoting by v the total velocity, so that

$$v^2 = \frac{\partial\phi}{\partial X_\beta}\frac{\partial\phi}{\partial X_\beta} , \tag{4.14}$$

the Hamilton-Jacobi equation (4.10) gives

$$X\frac{\partial v_r}{\partial s} + v^2 + K_{\alpha\beta}X_\alpha X_\beta = 0 . \tag{4.15}$$

[14] Ref. 13, p. 315.

This relation connects distance, radial velocity, total velocity and potential energy in the cloud of points corresponding to geodesics emanating from a point on g_0.

We shall use the above results to prove the following

THEOREM VIII. *If a congruence of geodesics emanates from a point* $s = 0$ *on a geodesic* g_0 *along which the principal curvatures are greater than or equal to a positive constant* K_0, *one of these geodesics intersects* g_0 *again to the first order at a point for which* $s \leqq \pi K_0^{-1/2}$.

As shown by Schoenberg (ref. 2, p. 487) the above theorem can be deduced immediately from Theorem I and a result established by Hadamard.[15] Hadamard's method involves a consideration of the second variation and is quite different from the present method, which appears to give a simpler and more geometrical approach to the question under consideration. The present theorem also follows (Myers, ref. 1) from the comparison theorem of Morse.[16]

To prove the theorem, we consider in the normal vector space the cloud of points corresponding to the emanation of a congruence of geodesics from g_0 at $s = 0$, at all inclinations up to a small angle $\bar{\theta}$.

It follows from the reasoning leading to (4.11) that v_r/X is the same for all points on a radius vector drawn from O, s being assigned. Thus v_r/X is a function only of the direction ratios of this line and of s. We define uniquely a function χ of these same variables by

$$\tan \chi = v_r/XK_0^{1/2}, \qquad -\tfrac{1}{2}\pi \leqq \chi \leqq \tfrac{1}{2}\pi. \tag{4.16}$$

If X is held fixed, χ varies in the same sense as v_r. We note that

$$\lim_{s\to 0} v_r/X = +\infty, \qquad \lim_{s\to 0} \chi = \tfrac{1}{2}\pi. \tag{4.17}$$

Now by the condition stated in the enunciation and by (2.13) we have

$$K_{\alpha\beta}X_\alpha X_\beta \geqq K_0 X_\alpha X_\alpha \tag{4.18}$$

for arbitrary X_α, and also obviously

$$v^2 \geqq v_r^2. \tag{4.19}$$

Consequently (4.15) gives

$$X\frac{\partial v_r}{\partial s} \leqq -v_r^2 - K_0X^2. \tag{4.20}$$

Confining our attention to a fixed point X_α, we have, by (4.16),

$$d\left(\frac{v_r}{XK_0^{1/2}}\right) \leqq -\frac{v_r^2 + K_0X^2}{X^2K_0^{1/2}}\,ds, \qquad d\chi \leqq -K_0^{1/2}\,ds, \tag{4.21}$$

and hence, by (4.17),

[15] J. Hadamard, *Leçons sur le Calcul des Variations*, vol. 1, Paris, 1910, p. 354.
[16] M. Morse, Math. Annalen, vol. 103 (1930), p. 66.

$$(4.22) \qquad \chi - \tfrac{1}{2}\pi \leqq - sK_0^{1/2}, \qquad v_r/XK_0^{1/2} \leqq \cot sK_0^{1/2}.$$

This result will hold provided the point under consideration has been immersed in the cloud for all values of s satisfying (4.23) below.

We shall call any positive value of s corresponding to an intersection to the first order of one of the geodesics with g_0 *a critical value*. Until such time as s passes through a critical value, all points in the immediate neighborhood of the origin remain immersed in the cloud. Hence, assuming that there is no critical value of s in the range

$$(4.23) \qquad 0 < s < \pi K_0^{-1/2},$$

the inequality (4.22) holds for this range. But as s approaches $\pi K_0^{-1/2}$, the right hand side of (4.22) approaches $-\infty$, and so, consequently, does the left hand side. But it is obvious that the equations (4.1) are such that v (and consequently v_r) is finite for every finite s. Hence X tends to zero as s tends to $\pi K_0^{-1/2}$. Thus we get an intersection either in the range (4.23) (invalidating the inequality (4.22)) or at the point $s = \pi K_0^{-1/2}$, which proves Theorem VIII.

Incidentally, we may deduce from (4.22) the fact that v_r is negative, and consequently all the geodesics are approaching g_0, if s is greater than $\frac{1}{2}\pi K_0^{-1/2}$ and less than the smallest critical value.

UNIVERSITY OF TORONTO.

ON VOLTERRA-STIELTJES INTEGRAL EQUATIONS

By W. C. Randels

We propose to prove an existence theorem for integral equations of the type

$$f(x) = g(x) + \lambda \int_0^x f(y) d_y K(x, y) \,. \tag{1}$$

The integral used is the Young-Stieltjes integral over the open interval $(0, x)$. The theorem is no longer true if an integral over the closed interval is used. We assume that $g(x)$ is bounded and Borel measurable on $(0, 1)$ and $K(x, y)$ is subject to the conditions

(A) $K(x, y)$ is Borel measurable as a function of x.

(B) There exists a monotone increasing and bounded function $V(y)$ such that

$$| K(x, y_1) - K(x, y_2) | \leqq | V(y_1) - V(y_2) | \,.$$

(We shall assume that $V(0) = 0$).

(C)[1] $K(x, y) = K(x, y - 0)$.

(D)[2] $K(x, 0) = 0$.

A function $f(x)$ is said to be a solution of (1) if it is bounded and Borel measurable and satisfies (1).

The essential difference between this problem and that of ordinary Volterra integral equations is that integration by parts of Stieltjes integrals is not permissible, so that it is necessary to obtain a formula which will replace integration by parts.[3] We do this as a lemma.

Lemma. *If $f(x)$ is positive, bounded and Borel measurable and $g_1(x)$, $g_2(x)$ are monotone increasing, bounded and continuous on the left, then*

$$\int_0^1 f(x) d[g_1(x) g_2(x)] \geqq \int_0^1 f(x) g_1(x) dg_2(x) + \int_0^1 f(x) g_2(x) dg_1(x) \,. \tag{2}$$

Proof. We shall first prove the lemma when $g_1(x)$ and $g_2(x)$ are both step functions. We choose the set of points $\{x_n\}$ composed of all points of discontinuity of $g_1(x)$ and $g_2(x)$. Since the functions are continuous on the left, we have

Received May 2, 1935. The author is a Sterling Research Fellow.

[1] The condition (C) entails no loss of generality, for the value of a function $h(x)$ of bounded variation may be changed at every point of discontinuity without changing $\int_0^1 f(x) dh(x)$.

[2] The condition (D) clearly involves no loss of generality.

[3] When $K(x, y)$ is continuous in y, integration by parts is permissible. This case of the problem has been solved by Tamarkin. See abstract, Bull. Amer. Math. Soc., vol. 35 (1929), p. 165.

$$
\begin{aligned}
&g_1(x_n+0)g_2(x_n+0)-g_1(x_n-0)g_2(x_n-0)\\
&\quad=\tfrac{1}{2}[g_1(x_n+0)+g_1(x_n-0)]\,[g_2(x_n+0)-g_2(x_n-0)]\\
&\quad+\tfrac{1}{2}[g_2(x_n+0)+g_2(x_n-0)]\,[g_1(x_n+0)-g_1(x_n-0)]\\
&\quad\geqq g_1(x_n)[g_2(x_n+0)-g_2(x_n-0)]+g_2(x_n)[g_1(x_n+0)-g_1(x_n-0)]\,,
\end{aligned}
$$

and

$$
\begin{aligned}
&\int_0^1 f(x)d[g_1(x)g_2(x)]=\lim\sum f(x_n)[g_1(x_n+0)g_2(x_n+0)-g_1(x_n-0)g_2(x_n-0)]\\
&\quad\geqq \lim\sum f(x_n)g_1(x_n)[g_2(x_n+0)-g_2(x_n-0)]\\
&\quad+\lim\sum f(x_n)g_2(x_n)[g_1(x_n+0)-g_1(x_n-0)]\\
&\quad=\int_0^1 f(x)g_1(x)dg_2(x)+\int_0^1 f(x)g_2(x)dg_1(x)\,.
\end{aligned}
$$

Now if $f(x)$ is continuous, by following through an argument of Bray[4] it can be easily seen that, if either $g_1(x)$ or $g_2(x)$ is continuous, then

$$
\int_0^1 f(x)d[g_1(x)g_2(x)]=\int_0^1 f(x)g_1(x)dg_2(x)+\int_0^1 f(x)g_2(x)dg_1(x)\,, \tag{3}
$$

and therefore by transfinite induction (3) holds for all bounded and Borel measurable $f(x)$. We know, however, that every function $g(x)$ of bounded variation can be expressed as the sum $g(x) = g'(x) + g''(x)$, where $g'(x)$ is continuous and $g''(x)$ is a step function. Therefore

$$
\begin{aligned}
&\int_0^1 f(x)d[g_1(x)g_2(x)]\\
&\quad=\int_0^1 f(x)d[g_1'(x)g_2'(x)+g_1'(x)g_2''(x)+g_1''(x)g_2'(x)+g_1''(x)g_2''(x)]\\
&\quad\geqq\int_0^1 f(x)g_1(x)dg_2(x)+\int_0^1 f(x)g_2(x)dg_1(x)\,.
\end{aligned}
$$

This completes the proof of the lemma.

It is now necessary to consider some properties of the function

$$
\bar{f}(x)=\int_0^x f(y)d_yK(x,y)\,,
$$

where $f(x)$ is bounded and Borel measurable and $K(x, y)$ satisfies (A), (B), (C) and (D). For convenience we extend the definition of $K(x, y)$ by setting

$$
K(x,y)=K(x,x)=K(x,x-0) \qquad y\geqq x\,,
$$

[4] H. E. Bray, Ann. of Math., (2), vol. 20 (1919), pp. 177–186. See Theorem 6.

and we have

$$\bar{f}(x) = \int_0^1 f(y) d_y K(x, y) .$$

We first notice that

$$\left| \int_0^1 f(y) d_y K(x, y) \right| \leqq \underset{0 < y < 1}{\text{l.u.b.}} |f(y)| \cdot V(1) ,$$

so that $\bar{f}(x)$ is bounded on $(0, 1)$. We now consider the special case where

$$f(x) = \begin{cases} 1, & 0 \leqq a < x < b \leqq 1 , \\ 0, & 0 \leqq x \leqq a , \quad b \leqq x \leqq 1 . \end{cases}$$

By the definition of the Young-Stieltjes integral, in this case

$$\bar{f}(x) = \int_0^1 f(y) d_y K(x, y) = \lim_{\substack{y_1 \to a+0 \\ y_2 \to b-0}} [K(x, y_2) - K(x, y_1)] .$$

But by (A), we know that for a fixed y_1 and y_2 the function $K(x, y_2) - K(x, y_1)$ is Borel measurable, and therefore $\bar{f}(x)$ being the limit of Borel measurable functions is itself Borel measurable. Moreover, every Borel measurable function may be obtained by a sequence, perhaps transfinite, of limiting operations starting from the set of step functions. Therefore, by transfinite induction for every bounded and Borel measurable $f(x)$, the function $\bar{f}(x)$ is bounded and Borel measurable.[5]

It is therefore possible to construct the sequence of kernels

$$K^{(n)}(x, y) = \int_0^1 K^{(n-1)}(z, y) d_z K(x, z) , \qquad K^{(1)}(x, y) = K(x, y) ,$$

and we have

$$K^{(n)}(x, y) = \int_0^x d_{z_1} K(x, z_1) \int_0^{z_1} d_{z_2} K(z_1, z_2) \cdots \int_0^{z_{n-2}} d_{z_{n-1}} K(z_{n-2}, z_{n-1}) \, K(z_{n-1}, y) ,$$

$$| K^{(n)}(x, y) | \leqq \int_0^x d_{z_1} V(z_1) \cdots \int_0^{z_{n-2}} d_{z_{n-1}} V(z_{n-1}) \, V(y) .$$

We may suppose that $V(y) = V(y - 0)$, since, if $y_2 > y_1$,

$$| K(x, y_2) - K(x, y_1) | = \lim_{y \to y_2 - 0} | K(x, y) - K(x, y_1) |$$

$$\leqq \lim_{y \to y_2 - 0} [V(y) - V(y_1)] = V(y_2 - 0) - V(y_1) \leqq V(y_2 - 0) - V(y_1 - 0) .$$

[5] A function $f(x)$ can only be obtained from the step functions which we have considered if $f(0) = f(1) = 0$, but, as we have mentioned, the integral we consider is only over the open interval $(0, 1)$, and hence the value of $f(x)$ at 0 and 1 does not affect $\bar{f}(x)$.

Therefore by our lemma

$$\{V(x)\}^n = \int_0^x d\{V(x)\}^n \geqq \int_0^x V(x)d\{V(x)\}^{n-1} + \int_0^x \{V(x)\}^{n-1}dV(x)$$

$$\geqq n \int_0^x \{V(x)\}^{n-1}\, dV(x)\,,$$

and

$$\int_0^x dV(z_1) \cdots \int_0^{z_{n-2}} dV(z_{n-1}) \leqq \frac{\{V(x)\}^{n-1}}{(n-1)!}\,.$$

Hence

$$| K^{(n)}(x, y) | \leqq \frac{\{V(1)\}^{n-1}\, V(y)}{(n-1)!}\,,$$

$$| K^{(n)}(x, y_2) - K^{(n)}(x, y_1) | \leqq \frac{\{V(1)\}^{n-1}\, | V(y_2) - V(y_1) |}{(n-1)!}\,.$$

Consequently the kernel

$$R(x, y; \lambda) = \sum_{n=1}^{\infty} \lambda^n K^{(n)}(x, y)$$

is defined for all values of λ and

$$| R(x, y_2; \lambda) - R(x, y_1; \lambda) | \leqq | V(y_2) - V(y_1) | \sum_{n=1}^{\infty} \frac{\lambda^n \{V(1)\}^n}{(n-1)!}\,,$$

so that $R(x, y; \lambda)$ is of bounded variation as a function of y. Since the functions $K^{(n)}(x, y)$ are Borel measurable as functions of x, $R(x, y; \lambda)$ is also.

We now assert that the function

$$f(x) = g(x) + \int_0^x g(y)d_yR(x, y; \lambda)$$

is a solution of (1). It is clear that $f(x)$ is bounded and Borel measurable. Now let us form

$$\lambda \int_0^x f(y)d_yK(x, y) = \lambda \int_0^1 g(y)d_yK(x, y) + \lambda \int_0^1 d_yK(x, y) \int_0^1 g(z)d_zR(y, z; \lambda)\,.$$

By the Fubini theorem

$$\int_0^1 d_yK(x, y)\Big[\sum_{n=1}^{\infty} \lambda^n \int_0^1 g(z)d_zK^{(n)}(y, z)\Big]$$

$$= \sum_{n=1}^{\infty} \lambda^n \int_0^1 g(z)d_z\Big[\int_0^1 K^{(n)}(y, z)d_yK(x, y)\Big]$$

$$= \sum_{n=1}^{\infty} \lambda^n \int_0^1 g(z)d_zK^{(n+1)}(x, z),$$

or

$$\lambda \int_0^x f(y)d_yK(x, y) = \sum_{n=1}^{\infty} \lambda^n \int_0^1 g(y)d_yK^{(n)}(x, y) = f(x) - g(x).$$

This proves that $f(x)$ is a solution of (1). The solution is unique, for, if there were two solutions, there would have to be a function $f(x)$ different from zero such that

$$f(x) = \lambda \int_0^x f(y) d_y K(x, y),$$

but then

$$\begin{aligned}
0 &= f(x) - \lambda \int_0^x f(y) d_y K(x, y) \\
&= \int_0^x f(z) d_z R(x, z; \lambda) - \lambda \int_0^x d_z R(x, z; \lambda) \int_0^z f(y) d_y K(z, y) \\
&= \int_0^x f(z) d_z R(x, z; \lambda) - \lambda \int_0^x f(y) d_y \int_0^x K(z, y) d_z R(x, z; \lambda) \\
&= \int_0^x f(z) d_z R(x, z; \lambda) - \int_0^x f(y) d_y R(x, y; \lambda) + \lambda \int_0^x f(y) d_y K(x, y) = f(x).
\end{aligned}$$

It is of some interest to notice that the condition (B) cannot be replaced by a condition of the form

$$\text{(B}'\text{)} \qquad \int_0^1 | d_y K(x, y) | < M.$$

To show this, we consider the kernel

$$K(x, y) = \begin{cases} 0, & y \leqq x/2, \\ 1, & y > x/2. \end{cases}$$

The homogeneous equation for this kernel for $\lambda = 1$ is $f(x) = f(x/2)$ and we see that if $f(x)$ is defined arbitrarily on the interval $\frac{1}{2} < x \leqq 1$, we can determine a solution $f(x)$ of the homogeneous equation. Furthermore, the equation (1) for $\lambda = 1$ is

$$f(x) = g(x) + f(x/2),$$

or

$$f(x/2^j) = f(x) - \sum_{i=0}^{j-1} g(x/2^i),$$

and hence $f(x)$ is not bounded unless $\sum_{i=0}^{N} g(x/2^i)$ is bounded for every N and every x on the interval $(\frac{1}{2}, 1)$. Since this need not be true, we see that, in some cases at least, the equation (1) will not have a solution in the sense that we have defined.

YALE UNIVERSITY.

ON LOCALLY CONNECTED SPACES

By R. L. Wilder

In a recent paper,[1] for the purpose of characterizing the boundaries of uniformly locally i-connected domains in n-space, I have designated a certain class of compact metric spaces as "generalized closed n-manifolds" (= g. c. n-m.). Considered as boundaries of domains in euclidean spaces, they form the exact analogues, for higher dimensions, of the domain boundaries in 3-space represented by the class of all closed 2-dimensional orientable manifolds; in particular, their Betti groups are finite and they satisfy the Poincaré duality, and, moreover, have the same sort of relations (as regards linkings, dualities, etc.) to their complements as have the ordinary 2-manifolds to their boundaries in 3-space; in 3-space they are identically the ordinary 2-manifolds. Among the problems so far not treated is that of proving the finiteness of the Betti numbers of the g. c. n-m. when considered as an abstract space, not necessarily imbedded in euclidean space. It was in considering this problem that the results of the present paper were obtained.

From the properties of the abstract g. c. n-m. it follows that such a space is locally i-connected, in the sense of Vietoris cycles (= V-cycles),[2] for $0 \leqq i \leqq n-2$.

Received May 22, 1935.

[1] *Generalized closed manifolds in n-space*, Annals of Mathematics, vol. 35 (1934), pp. 876–903. We refer to this paper hereafter as G. C. M.

[2] Shortly after this paper was presented to the American Mathematical Society (see its Bulletin, vol. 41 (1935), p. 202, abstract no. 178), there appeared two papers intimately related, in part, to the subject matter of the present paper. We refer to the following: K. Borsuk, *Un théorème sur les groupes de Betti des ensembles localement connexes en toutes les dimensions* $\leqq n$, Fundamenta Mathematicae, vol. 24 (1935), pp. 311–316; S. Lefschetz, *Chain-deformations in topology*, this Journal, vol. 1 (1935), pp. 1–18. In the former article, the finiteness of the Betti groups is established for spaces locally connected in the sense of Lefschetz's *Topology* (p. 91); in the latter, a similar result is obtained using a more general type of local connectedness. The reason for our use of Vietoris cycles, which we term V-cycles hereafter, in the local connectedness is due to the origin of our problem in the g. c. n-m., whose origin in turn in the study of domain boundaries necessitated a formulation of their properties in terms of such cycles. As a point of departure, it seems to us that the local connectedness in the sense of Lefschetz may be preferable in that it avoids certain complications of proof encountered in using the "infinite" cycles; thus, in the present work we have had to restrict our chains to those obtainable using a finite coefficient ring, for reasons of convergence, whereas in using singular chains the type of coefficient ring used is not material. Apparently, however, local connectedness over a range of dimensions $0 \leqq i \leqq k$ is more general in the sense of the V-cycles than in the sense of either singular spheres or cycles, since the two latter types imply the former but not conversely. For instance, we may construct a space consisting of an infinite set of mutually exclusive "Poincaré spaces" converging to a point and successively connected by simple arcs which

Consequently, if we can show that a compact metric space which is locally i-connected for $0 \leqq i \leqq k$ (k a non-negative integer) has finite corresponding i-dimensional Betti numbers, we shall have as a corollary that the numbers $p^i(M)$ of a g. c. n-m. M are finite for $0 \leqq i \leqq n-2$; the finiteness of $p^{n-1}(M)$ is a consequence of a further property of the g. c. n-m., namely, what we have called below its *semi-$(n-1)$-connectedness*. As a by-product of our investigations we have considered spaces locally connected for dimensions $\leqq k$ as generalizations of the Jordan (or Peano) continua, and obtained an n-space generalization of the well-known Schoenflies-Torhorst theorem concerning the jordanian character of the boundaries of the domains complementary to plane Jordan continua.

Because the methods we employ depend upon many successive deletions of terms from a given infinite series, we shall find it convenient to introduce the following

Definition. Let S be a sequence $e_{s_1}, e_{s_2}, \cdots, e_{s_k}, \cdots$ of symbols, where the subscripts s_k are natural numbers, and let $n_1, n_2, \cdots, n_m, \cdots$ be a subsequence, N, of the sequence of subscripts s_k. Then we say that the sequence S is *harmonized* with the sequence N if from S we delete those symbols whose subscripts are not in N. In the resulting subsequence, S', of S, the relative order of the terms is the same as in the original sequence S. We shall also say that S' is *in harmony with* N.

We deal, at first, with a compact metric space M, and we recall, once and for all, that if $\gamma^i = \{i_1, i_2, \cdots, i_k, \cdots\}$ is a V-cycle of M, and N an infinite sequence $n_1, n_2, n_3, \cdots$, of natural numbers monotonically increasing, then the cycle $\Gamma^i = \{i_{n_1}, i_{n_2}, \cdots, i_{n_k}, \cdots\}$ is equivalent on M, in the homology sense, to γ^i. In other words, if the sequence constituting a V-cycle is harmonized with an infinite subsequence of its subscripts, the resulting cycle will be equivalent, in all our operations, to the original cycle. Our chains and cycles will have a finite coefficient ring, the integers mod $m \geqq 2$.

Suppose we have, in M, a sequence of abstract ϵ_k-chains

$$(1) \qquad K^{i+1} = \{K_1, K_2, \cdots, K_k, \cdots\},$$

where $\lim_{k\to\infty} \epsilon_k = 0$ and $K_k \to i_k$. We then call K^{i+1} a *V-chain* to distinguish it from an ordinary finite abstract chain such as, for instance, an individual K_k, and write $K^{i+1} \to \gamma^i$. In general, we shall write for the sake of brevity, $K^{i+1} =$

is locally 0-connected in any sense, and is locally 1-connected in the sense of the V-cycles, but is not locally 1-connected in the sense of *Topology*. However, for an isolated dimension i, the various senses of local connectedness are independent; thus, in the plane it is easy to construct a non-jordanian continuum which is locally 1-connected in the sense of singular spheres or cycles but not in terms of V-cycles. In closing these incidental remarks we point out that the finiteness of the Betti groups for spaces that are locally connected in all dimensions zero to infinity, according to the definition of *Topology*, was proved by Lefschetz in Annals of Mathematics, vol. 35 (1934), pp. 118–129. See also an earlier paper of Borsuk in Fundamenta Mathematicae, vol. 21 (1933), pp. 91–98.

$\{K_k\}$, meaning thereby (1). If we have another V-chain, say $H^{i+1} = \{H_k\}$, such that $H^{i+1} \to \gamma^i$, we may formally combine relations and write $K^{i+1} - H^{i+1} = \{K_k - H_k\} \to 0$. However, $K^{i+1} - H^{i+1}$ will not, in general, be a V-cycle. But by a convergence theorem of Alexandroff,[3] there exists a subsequence of the sequence $\{K_k - H_k\}$ which is a V-cycle. In our new terminology, there exists a sequence N of natural numbers n_k such that if $\{K_k - H_k\}$ is harmonized with N, the resulting sequence is a V-cycle. And if, moreover, we harmonize each of the sequences $\{i_k\}$, $\{K_k\}$, $\{H_k\}$ with N, we can now write formally, as in the combinatorial theory of finite complexes,

$$K^{i+1} \to \gamma^i\,; \qquad H^{i+1} \to \gamma^i\,; \qquad K^{i+1} - H^{i+1} \to 0\,,$$

with $K^{i+1} - H^{i+1}$ a V-cycle. It will be noted that we have, here, retained the same symbol for a V-chain or V-cycle after it has been harmonized, and we shall continue to do this, without further comment on the fact, in all that follows.

Now let M be locally i-connected for $0 \leqq i \leqq j$, where j is some fixed non-negative (usually positive) natural number, and let C^i be an abstract i-simplex of M: $C^i = a_0a_1 \cdots a_i$, the a's being the symbols for the "vertices". We shall say that C^i has a *V-chain-realization* or simply a *chain-realization* in M if the following procedure can be carried out.[4] Each pair of vertices a_h, a_m $(h < m)$ forms a 0-cycle which we suppose bounds a V-chain $K^1_{hm} = \{1^k_{hm}\}$, where $1^k_{hm} \to a_h - a_m$ is an ϵ_k-chain, $\lim_{k\to\infty} \epsilon_k = 0$. We make the convention, and like conventions in following cases, that $-K^1_{hm} = \{-1^k_{hm}\}$ is the opposite orientation of K^1_{hm}. Now consider, for instance, the sequence $K^1_{01} - K^1_{02} + K^1_{12} = \{1^k_{01} - 1^k_{02} + 1^k_{12}\}$; this has a convergent subsequence forming a V-cycle $\gamma^1_{012} = \{1^{n_k}_{01} - 1^{n_k}_{02} + 1^{n_k}_{12}\}$. We now harmonize the preceding sequences with the sequence of natural numbers n_k. The sequence $K^1_{01} - K^1_{03} + K^1_{13} = \{1^{n_k}_{01} - 1^{n_k}_{03} + 1^{n_k}_{13}\}$ has a convergent subsequence $\gamma^1_{013} = \{1^{r_k}_{01} - 1^{r_k}_{03} + 1^{r_k}_{13}\}$, and all preceding sequences are harmonized with the sequence $\{r_k\}$. At the end of the second stage of our procedure, we have a set of sequences K^1_{hm}, γ^1_{hms} all harmonized with the last set of subscripts m_k obtained in getting a convergent sequence, and each term of every sequence consisting of cycles on the a's.

In the third stage of the procedure, we suppose each γ^1_{hms} bounds a V-chain $K^2_{hms} = \{2^{m_k}_{hms}\}$. Then, starting with the sequence

$$\{K^2_{123} - K^2_{023} + K^2_{013} - K^2_{012}\}\,,$$

we choose a convergent subsequence $\gamma^2_{0123} = \{2^{t_k}_{123} - 2^{t_k}_{023} + 2^{t_k}_{013} - 2^{t_k}_{012}\}$, and then harmonize all above-mentioned sequences with the sequence of numbers t_k, and so on.

[3] P. Alexandroff, *Dimensionstheorie*, Mathematische Annalen, vol. 106 (1932), pp. 161–238; see pp. 180–181 for definition of convergent cycle and the convergence theorem.

[4] Compare the filling-in process used by Lefschetz in the papers cited above.

Finally, we obtain a single $(i-1)$-cycle $\gamma^{i-1}_{012\cdots i}$ which, we suppose, bounds a sequence $K^i = K^i_{01\cdots i} = \{i^{u_k}_{01\cdots i}\}$. The chain-realization of C^i consists of all sequences $K^\lambda_{hm\cdots r}$ and V-cycles $\gamma^{t-1}_{hm\cdots s}$ harmonized with the sequence $\{u_k\}$. The sequences $K^\lambda_{hm\cdots s}$, properly oriented, form the chain-realizations of the lower dimensional cells or faces of C^i.

The extension of the above process to obtain chain-realizations of an abstract i-complex is made in an obvious manner. There being in this case more than one i-cell involved, in general, it is necessary to obtain more than one sequence K^i, although it will not be necessary to harmonize any sequences in this (final) step of the process.

Consider now chain-realizations of i-chains. If K is an abstract complex of M, let $L^i = c_1E^i_1 + \cdots + c_hE^i_h$ be a chain based on the i-cells

$$E^i_m = a^m_0 a^m_1 \cdots a^m_i$$

of K $(m = 1, 2, \cdots, h)$. We get chain-realizations of the cells of K as above, properly oriented, and let the chain-realization of E^i_m be denoted by $K^i_m = \{e^k_m\}$. Then $c_mE^i_m$ is "realized" by $c_mK^i_m = \{c_me^k_m\}$, and the chain-realization of L^i is taken to be $\bar{L}^i = c_1K^i_1 + \cdots + c_hK^i_h$.

The following lemma is easily established; see, for instance, the proof of Lemma 2 of G. C. M.

LEMMA 1. *If M is a locally i-connected $(0 \leqq i \leqq j)$, compact metric space, then for any $\epsilon > 0$ there exists a $\delta > 0$ such that any abstract λ-cell $(0 \leqq \lambda \leqq j+1)$ of M of diameter $< \delta$ has a V-chain-realization of diameter $< \epsilon$.*[5]

We now make a distinction between the usual V-cycle and the more specialized type of cycle which we shall call "constructible".

DEFINITION. A convergent cycle $\gamma^i = \{i_k\}$ of M will be called *constructible* if for some value of k, say m, the cycle $\bar{\gamma}^i = \{i_m, i_{m+1}, \cdots\}$ is the chain-realization of the abstract cycle i_m.

Although we prove directly a more general theorem, we first prove, in order to separate the difficulties, the following

THEOREM 1. *The Betti numbers $p^i(M)$ of a locally i-connected compact metric space, $0 \leqq i \leqq j$, are all finite.*

Proof. Vietoris has shown[6] that there exists in M a basis $F_1, F_2, \cdots, F_k, \cdots$ of convergent i-cycles such that any $\gamma^i \sim c_1F_1 + \cdots + c_kF_k + \cdots$. This basis has the property that its first n_1 members form an ϵ_1-basis for M, whereas for $k > n_1$, $F_k \underset{\epsilon_1}{\sim} 0$; its first n_2 members form an ϵ_2-basis for M, whereas for $k > n_2$, $F_k \underset{\epsilon_2}{\sim} 0$; and so on. We now replace this basis by a basis of constructible cycles, in the following manner. Let $F_s = \{i^s_k\}$. For $n_{h-1} < s \leqq n_h$,

[5] By the diameter of a chain or chain-realization we mean the diameter of the minimal closed point set that carries it.

[6] L. Vietoris, *Über den höheren Zusammenhang kompakter Räume und eine Klasse von zusammenhangstreuen Abbildungen*, Mathematische Annalen, vol. 97 (1927), pp. 454–472. This paper contains the original definition of the V-cycles (*Fundamentalfolgen*), bases, connectivity numbers, etc.

we may first assure ourselves that all elements i_k^s are already $\underset{\epsilon_{h-1}}{\sim} 0$, and that for any subscripts, $m, l, i_m^s \underset{\epsilon_h}{\sim} i_l^s$.

We next determine a δ_h such that any i-cell of M of diameter $< \delta_h$ has a chain-realization of diameter $< \epsilon_h$. For any fixed s, we determine $k = \bar{k}$ such that $i_{\bar{k}}^s$ is a δ_h-chain of M—that is, each of its fundamental cells is of diameter $< \delta_h$. By the above described process we get a chain-realization of $i_{\bar{k}}^s$ such that each chain-realization of a fundamental i-cell is in an ϵ_h-neighborhood of that cell, and contains the vertices of that cell. Thus, if

$$i_{\bar{k}}^s = c_1 C_1^i + c_2 C_2^i + \cdots + c_g C_g^i ,$$

where the C^i's are the abstract i-cells, we get a chain-realization

$$\begin{aligned} \gamma_s^i &= c_1 K_1^i + c_2 K_2^i + \cdots + c_g K_g^i \\ &= \{c_1\, {}_1 i^{u_k}_{0\,1\cdots i} + \cdots + c_g\, {}_g i^{u_k}_{0\,1\cdots i}\} . \end{aligned}$$

The vertices of each C^i are contained among those of the corresponding $i^{u_k}_{0\,1\cdots i}$'s, the entire corresponding K^i lying in an ϵ_h-neighborhood of C^i. Consequently, by keeping the vertices belonging to the C^i's fixed, and moving the other vertices of $i^{u_k}_{0\,1\cdots i}$ arbitrarily into those of the corresponding C^i, we get a simplicial ϵ_h-mapping of the elements of the chain γ_s^i into $i_{\bar{k}}^s$, so that

$$i_k^s \underset{\epsilon_h}{\sim} c_1\, {}_1 i^{u_k}_{0\,1\cdots h} + \cdots + c_g\, {}_g i^{u_k}_{0\,1\cdots i}$$

for all k.

The constructible cycle

$$\Gamma_s^i = \{i_1^s, \cdots, i_{\bar{k}}^s, c_1\, {}_1 i^{u_1}_{0\,1\cdots i} + \cdots + c_g\, {}_g i^{u_1}_{0\,1\cdots i}, \cdots\}$$

is ϵ_h-homologous to F_s on M, and since $\epsilon_h < \epsilon_{h-1}$, from $F_s \underset{\epsilon_{h-1}}{\sim} 0$ follows $\Gamma_s^i \underset{\epsilon_{h-1}}{\sim} 0$ on M.

To show that the number $p^i(M)$ is finite, we select a number N so great that for $s > N$, $n_{h-1} < s \leqq n_h$, the corresponding $\epsilon_{h-1} < \delta$, where δ is such that any δ-cell of M has a chain-realization in M.[7] Then, for any such s, $\Gamma_s^i \underset{\epsilon_{h-1}}{\sim} 0$ on M, and, in particular, there exists an abstract ϵ_{h-1}-chain $K_s^i \rightarrow i_{\bar{k}}^s$ in M. Now the chain K_s^i can be given a chain-realization in M along with the construction of Γ_s^i, which was constructed on $i_{\bar{k}}^s$. Consequently $\Gamma_s^i \sim 0$ in M, and therefore $\Gamma_s^i \underset{\epsilon_h}{\sim} 0$ in M. But we already have that $F_s \underset{\epsilon_h}{\sim} \Gamma_s^i$ in M and hence it follows that $F_s \underset{\epsilon_h}{\sim} 0$ in M, contradicting the fact that F_s is in the ϵ_h-basis of M. It follows that the basis of F's is finite.

We turn now to metric spaces that are locally compact.

Lemma 2. *Let F be a self-compact subset of the locally i-connected* $(0 \leqq i \leqq j)$, *locally compact metric space M. Also, let U be an open subset of M containing F. Then for any $\epsilon > 0$ there exists a $\delta > 0$ such that any abstract λ-cell of F of*

[7] We may assume that the δ's form a monotonically decreasing sequence.

diameter $< \delta$ *has a V-chain-realization of diameter* $< \epsilon$ *in U provided that* $0 \leqq \lambda \leqq j + 1$.

Proof. The lemma is true if for each λ there exists a $\delta_\lambda > 0$ such that abstract λ-cells of F of diameter $< \delta_\lambda$ have chain-realizations of diameter $< \epsilon$ in U. Accordingly, we prove the latter, which, being trivial for $\lambda = 0$, we may assume true for the dimension $\lambda - 1$ and prove for λ.

Let V be an open subset of M such that $U \supset V \supset F$, and $\bar{V}$, the closure of V, is self-compact. We may assume $\epsilon < \rho(\bar{V}, M - U)$. As $\bar{V}$ is self-compact, there exists $\delta' > 0$ such that any $(\lambda - 1)$-cycle of $\bar{V}$ of diameter $< \delta'$ bounds a chain of M of diameter $< \epsilon$; obviously such a chain must also be a chain of U.

By our assumption there exists $\delta > 0$ such that a $(\lambda - 1)$-cell of F of diameter $< \delta$ has a $\delta'/4$-chain-realization on V.

Let C^λ be an abstract cell of F of diameter $< \delta$. Then, as the $(\lambda - 1)$-cells of C^λ have $\delta'/4$-chain-realizations on V, the boundary $(\lambda - 1)$-chain of C^λ has a δ'-chain-realization on V; this is a cycle $\Gamma^{\lambda-1}$ of V of diameter $< \delta'$. By the definition of δ', there is a chain K^λ of U of diameter $< \epsilon$, bounded by $\Gamma^{\lambda-1}$. The chain K^λ is the desired realization of C^λ.

COROLLARY. *Under the hypothesis of Lemma* 2, *there exists* $\delta > 0$ *such that abstract* δ*-chains of F are chain-realizable in U.*

THEOREM 2. *Let F be a self-compact subset of the locally i-connected* $(0 \leqq i \leqq j)$, *locally compact metric space M. Also, let U be an open subset of M containing F. Then at most a finite number of i-dimensional V-cycles of F are independent with respect to homologies in U.*

Proof. Let V be an open subset of M such that $U \supset V \supset F$ and $\bar{V}$ is self-compact. By the corollary to Lemma 2 there exists a positive number δ_V such that abstract i-cells of $\bar{V}$ are chain-realizable in U, if of diameter $< \delta_V$.

Using the methods of Vietoris[8] we obtain a basis of i-cycles, F_s^i, $s = 1, 2, 3, \cdots$, of F for homologies in U which we suppose infinite; in what follows we preserve the Vietoris notation as employed in the proof of Theorem 1. By Lemma 2, there exists for each ϵ_h a δ_h such that any i-cell of F of diameter $< \delta_h$ has a chain-realization in V of diameter $< \epsilon_h$. We then proceed as in the proof of Theorem 1 to get constructible cycles $\Gamma_s^i \underset{\epsilon_h}{\approx} F_s$ on V. We have, from $F_s \underset{\epsilon_{h-1}}{\approx} 0$ on U, that $\Gamma_s^i \underset{\epsilon_{h-1}}{\approx} 0$ on U.

The cycles $\Gamma_{n_u} (u = 1, 2, 3, \cdots)$ are independent with respect to homologies in U. For suppose we have

$$c_1\Gamma_{n_{u_1}} + \cdots + c_m\Gamma_{n_{u_m}} + \cdots \sim 0 \text{ in } U; \qquad u_1 < \cdots < u_m < \cdots .$$

Then certainly

$$(2) \qquad c_1\Gamma_{n_{u_1}} + \cdots + c_m\Gamma_{n_{u_m}} + \cdots \underset{\epsilon_{u_1}}{\approx} 0 \text{ in } U.$$

But as $u_1 < u_k$ for $k > 1$, we have, by construction of the F's and Γ's, that

$$(3) \qquad c_k\Gamma_{n_{u_k}} \underset{\epsilon_{u_1}}{\approx} 0 \text{ in } U; \qquad k > 1.$$

[8] Loc. cit.

Combining relations (2) and (3), however, we get $c_1\Gamma_{n_{u_1}} \underset{\epsilon_{u_1}}{\sim} 0$ on U, contradicting its ϵ_{u_1}-independence with respect to homologies in U.

Let us now select a positive integer N such that for all $m > N$, $\delta_{u_m} < \frac{1}{3}\delta_V$, and consider the set of $\Gamma_{n_{u_m}}$'s such that $m > N$. The elements i_k^s corresponding to these Γ's (see proof of Theorem 1) cannot be linearly independent with respect to δ_V-homologies in $\bar{V}$.[9] There exists, indeed, an abstract chain K^{i+1} of $\bar{V}$ whose cells are of diameter $< \delta_V$ and whose boundary is a finite linear combination of the above-mentioned abstract cycles i_k^s. Then by chain-realizing in U the chain K^{i+1}, we simultaneously realize the Γ's corresponding to the i_k^s's on the boundary of K^{i+1}, and we thereby show that the Γ's are not independent with respect to homologies in U, contradicting the result of the preceding paragraph.

Since, in metric spaces that are only locally compact, local properties are not in general uniform over the entire space in terms of the given metric, we give the

DEFINITION. A metric space M will be called *semi-i-connected* if, given a point P of M, there exists an $\epsilon_P > 0$ such that all i-dimensional V-cycles of $S(P, \epsilon_p)$ bound on M.

It is trivial that every continuum is semi-0-connected; also that every locally i-connected space is semi-i-connected. However, the converses of these statements are not true. For compact spaces, where the ϵ may be determined uniformly for all points P, the property of semi-i-connectedness becomes merely the property that all "small" i-cycles bound.

LEMMA 3^i. *If the semi-i-connected, locally compact metric space M is not locally i-connected at the point P, there exists an $\epsilon > 0$ such that for any neighborhoods U, V and W of P for which $S(P, \epsilon) \supset U \supset V \supset W$, there exist on the boundary of V infinitely many i-dimensional V-cycles that are independent with respect to homologies in $\bar{U} - W$.*

Proof. We select $\epsilon > 0$ such that (1) all i-cycles of $S(P, \epsilon)$ bound on M, and (2) for any $\delta > 0$ there exists in $S(P, \delta)$ an i-cycle which does not bound in $S(P, \epsilon)$. Let U, V and W be neighborhoods of P such that $S(P, \epsilon) \supset U \supset V \supset W$. In W there exist infinitely many i-cycles that are independent with respect to homologies in U.[10] Consider any such cycle γ^i. We have $K^{i+1} \rightarrow \gamma^i$ on M. As K^{i+1} does not lie in $\overline{S(P, \epsilon)}$, it intersects $F(V)$, the boundary of V, so that on $F(V)$ there exists a cycle Γ^i such that $\Gamma^i \sim \gamma^i$ on $\bar{V}$. That the set of all such cycles Γ^i are independent with respect to homologies in $\bar{U} - W$ is easily shown.

THEOREM 3. *In order that a semi-i-connected, locally compact metric space should fail to be locally i-connected, it is necessary and sufficient that there exist concentric spheres $S(P, \epsilon)$, $S(P, \delta)$, and $S(P, \eta)$, where $\epsilon > \delta > \eta > 0$, such that on*

[9] As proved by Vietoris for the compact case; see his paper cited above, number (3').

[10] See P. Alexandroff, *On local properties of closed sets*, Annals of Mathematics, vol. 36 (1935), pp. 1–35; especially p. 9.

$F(P, \delta)$ there exist infinitely many i-dimensional V-cycles that are linearly independent with respect to homologies in $\overline{S(P, \epsilon)} - S(P, \eta)$.

The necessity of Theorem 3 follows from Lemma 3^i, and the sufficiency from Theorem 2. We remark that since every continuum is semi-0-connected, the characterization of non-locally-i-connected spaces contained in Theorem 3 was substantially obtained, for continua and $i = 0$, by R. L. Moore in the study of non-jordanian continua.[11] In Moore's theorem an infinite set of distinct components of $\overline{S(P, \epsilon)} - S(P, \eta)$ of the non-locally-0-connected M are obtained, all of which meet $F(P, \epsilon)$ and $F(P, \eta)$, and from whose intersections with $F(P, \delta)$ may be obtained the 0-cycles of our Theorem 3.

Lemma 4^i. *If the Betti number $p^i(M)$ of a compact metric space M is finite, then M is semi-i-connected.*

Proof. If a V-cycle γ^i fails to bound in M, there must exist a number $\epsilon > 0$ such that it is not $\underset{\epsilon}{\sim} 0$ on M. Consequently, since we are dealing with a finite coefficient ring, if $p^i(M)$ is finite, and thus M has a finite i-basis from which only a finite number of non-bounding linearly independent combinations can be formed, there exists an $\eta > 0$ such that no linear combination of its i-basis is $\underset{\eta}{\sim} 0$. Then any i-cycle of M of diameter $< \eta$ bounds on M. For otherwise, as it is homologous to, and hence η-homologous to a linear combination L of the elements of the i-basis, and is certainly itself $\underset{\eta}{\sim} 0$, we should have $L \underset{\eta}{\sim} 0$.

Theorem 4. *In order that a locally i-connected* ($0 \leqq i \leqq j$) *compact metric space M should have a finite Betti number $p^{j+1}(M)$, it is necessary and sufficient that it should be semi-$(j+1)$-connected.*

Proof. The necessity follows from Lemma 4^i with $i = j + 1$.

The sufficiency proof is as follows. As in the proof of Theorem 2, if $p^{j+1}(M)$ is infinite we may obtain a linearly independent infinite set of constructible $(j + 1)$-cycles $\Gamma_1, \Gamma_2, \Gamma_3, \cdots$, where $\Gamma_k \underset{\epsilon}{\sim} 0$ on M. By Lemma 1, there exists $\delta > 0$ such that any abstract $(j + 1)$-cell of M has an $\eta/3$-realization on M, where $\eta > 0$ is such that $(j + 1)$-cycles of M of diameter $< \eta$ bound on M.

Using notation as in the proof of Theorem 1, we may determine a Γ_s such that the abstract cycle i_k^s (where now $i = j + 1$) is a δ-cycle, and at the same time $\epsilon_s < \delta$. Then there exists on M an abstract ϵ_s-complex $K^{j+2} \to i_k^s$. We may now carry out, through cells of dimension $j + 1$, the simultaneous construction of K^{j+2} and i_k^s, where, due to the local i-connectedness of M for $0 \leqq i \leqq j$, the $(j + 1)$-cycles forming the chain-realizations of the $(j + 2)$-cell boundaries of K^{j+2} turn out to be of diameter $< \eta$. As such, these cycles bound $(j + 2)$-chains of M, which, although not "small", suffice to obtain a chain-realization of K^{j+2} whose boundary is Γ_s. Thus $\Gamma_s \sim 0$ on M, and this contradiction proves the theorem.

We next note that from Theorem 3 and the first half of condition (3) of the definition of g. c. n-m. in G. C. M., we have

[11] R. L. Moore, *Report on continuous curves from the viewpoint of analysis situs*, Bulletin of the American Mathematical Society, vol. 29 (1923), pp. 289–302, §3.

THEOREM 5. *A g. c. n-m. is locally i-connected for* $0 \leqq i \leqq n-2$.[12]

Combining the results of Theorems 1, 4 and 5, and the fact that by definition a g. c. n-m. is semi-$(n-1)$-connected, we have

THEOREM 6. *The Betti numbers of a g. c. n-m. are all finite.*

Jordan continua and generalizations

An interesting corollary of Theorem 4 from the standpoint of the theory of Jordan continua is the following

COROLLARY. *In order that the* 1-*dim. Betti number of a Jordan continuum should be finite, it is necessary and sufficient that it be semi*-1-*connected.*

A well-known theorem concerning plane Jordan continua[13] states that if M is a Jordan continuum in the plane and D is a domain complementary to M, then the boundary of D is itself a Jordan continuum. By use of the methods developed above, we can now prove a theorem true for all dimensions of which the result just stated is the case $n = 2$.[14]

THEOREM 7. *In* E_n,[15] $n \geqq 2$, *let* M *be a compact continuum which is locally i-connected for* $0 \leqq i \leqq n-2$. *Then, if* D *is a domain complementary to* M, *the boundary of* D *is a Jordan continuum.*[16]

Proof. Denote the boundary of D by B. Then B is a compact continuum and it is only necessary to show that it is locally 0-connected.[17]

First we show that the set $M' = E_n - D$ is locally i-connected for $0 \leqq i \leqq n-2$. That it is locally i-connected at interior points of M and at points of $E_n - M$ is obvious. We have, then, only frontier points of M to consider. Let P be such a point, and let ϵ be an arbitrary positive number. Let $\delta > 0$ be such that i-cycles of $M \cdot S(P, \delta)$ bound in $M \cdot S(P, \epsilon)$. Let $\gamma^i = \{i_k\}$ be a cycle of $M' \cdot S(P, \delta)$, and consider any V-chain $K^{i+1} \rightarrow \gamma^i$ in $S(P, \delta)$, where $K^{i+1} = \{L_k\}$. Let us fill in each cell of L_k geometrically. Then the set of those closed $(i+1)$-cells of L_k that do not meet M form a complex H_k. If H_k is always vacuous, then L_k may be assumed to lie on M, and no further proof

[12] We have not considered whether a g. c. n-m. is locally $(n-1)$-connected. It would be interesting to know if such is the case.

[13] See M. Torhorst, *Über den Rand der einfach zusammenhängenden ebenen Gebiete*, Mathematische Zeitschrift, vol. 9 (1921), pp. 44–65. See also my paper *Concerning continuous curves*, Fundamenta Mathematicae, vol. 7 (1925), pp. 340–377, Theorem 11.

[14] For an analogous theorem, relating, however, to absolute neighborhood retracts (which in E_n can have only finitely many complementary domains), see K. Borsuk, *Über eine Klasse von lokal zusammenhängenden Räumen*, Fundamenta Mathematicae, vol. 19 (1932), pp. 220–242, Satz 21. Also see G. C. M., Principal Theorem D and the remarks following it.

[15] We use the symbol E_n to denote euclidean n-space.

[16] It will be noted that we state nothing concerning the dimension of M, since if its dimension is less than $n-1$, the boundary of (the only possible) complementary domain is M itself and the result is trivial. We therefore may tacitly assume in giving the details of the proof that the dimension of M is $\geqq n-1$.

[17] By virtue of the well-known Hahn-Mazurkiewicz characterization of Jordan continua.

is necessary. Otherwise, the component $(i+1)$-complexes of H_k that contain points of $|i_k|$[18] form a complex C_k. If C_k is always vacuous, the i_k's may be assumed to lie on M, and again (because of the way δ was chosen) no further proof is necessary. The complex C_k lies wholly in M', since if a component of C_k has a point in D it must lie wholly in D and hence contain no point of $|i_k|$. We delete the cells of C_k from the complex consisting of cells in L_k, to form a complex N_k. By assigning to each $(i+1)$-cell of N_k that coefficient which it has in L_k, we obtain a new $(i+1)$-chain $K_k \to i'_k$. As we are dealing with a finite coefficient ring, we may assume the sequence $\Gamma^i = \{i'_k\}$ forms a V-cycle—i.e., we assume that the above process of harmonizing chains has been carried out. Then we have

$$(4) \qquad L_k - K_k \to i_k - i'_k, \qquad k = 1, 2, 3, \cdots$$

as the elements of a chain based only on cells of C_k, hence a chain in $M' \cdot S(P, \epsilon)$.

Now Γ^i may be considered as a cycle of M, infinitesimally removed,[19] and as such there exists

$$(5) \qquad Q_k \to i'_k \qquad \text{on } M \cdot S(P, \epsilon), \qquad k = 1, 2, 3, \cdots .$$

Combining relations (4) and (5) we get

$$L_k - K_k + Q_k \to i_k \qquad \text{on } M' \cdot S(P, \epsilon), \qquad k = 1, 2, 3, \cdots .$$

Consequently M' is locally i-connected.

Now suppose B is not locally 0-connected. Then $D + B$ is not locally 0-connected, as may be shown in a manner similar to that employed in the first paragraph of p. 887 of G. C. M., and there exist[20] concentric $(n-1)$-spheres K_1 and K_2 and a sequence of sub-continua of $D + B$, viz., $H_1, H_2, H_3, \cdots$, such that (1) each H_k lies in $K_1 + K_2 + I$ (where I is the "shell" domain bounded by $K_1 + K_2$) and contains at least one point of each of the spheres K_1, K_2; (2) no two of the continua H_k have a point in common, and moreover each is a component of $(D + B) \cdot (K_1 + K_2 + I)$.

Let K_3, K_4 and K_5 be $(n-1)$-spheres concentric with K_1 and K_2 and such that if r_j is the radius of K_j $(j = 1, 2, \cdots, 5)$,

$$r_1 > r_3 > r_4 > r_5 > r_2 .$$

For any given $\epsilon > 0$, there exist infinitely many distinct components of $D \cdot (K_1 + K_2 + I)$ which have points in $S(K_4, \epsilon)$. For suppose the contrary. Consider the sets H_k. A component of $D \cdot (K_1 + K_2 + I)$, if it has a point in common with a set H_k at all, is a subset of that H_k. Therefore, infinitely many sets $H_k \cdot S(K_4, \epsilon)$ contain no points of $D \cdot (K_1 + K_4 + I)$ and must as a

[18] If K is a complex, we denote the set of points on K by $|K|$.

[19] See P. Alexandroff, *Gestalt und Lage* . . . , Annals of Mathematics, vol. 30 (1928–29), pp. 101–187, Anhang I.

[20] R. L. Moore, loc. cit.

consequence be subsets of B—i.e., must consist entirely of limit points of $D\cdot(K_1 + K_2 + I)$. But then we should have different sets $H_k\cdot S(K_4, \epsilon)$ containing limit points of the same component of $D\cdot(K_1 + K_2 + I)$. This is impossible.

Let $\delta_1, \delta_2, \cdots, \delta_m, \cdots$ be a sequence of positive numbers monotonically decreasing to zero. Then there exist mutually exclusive components $D_1, D_2, \cdots, D_k, \cdots$ of $D\cdot(K_1 + K_2 + I)$ such that D_k contains a point of $S(K, \delta_k)$. For each k, let P_k be a point of D_k within a distance δ_k of K_4. Let s_k be an arc of D joining P_k to an arbitrary point of $D - D_k$. Such an arc must meet K_1 or K_2. It therefore has a subarc t_k joining P_k to a point Q_k either of K_1 or K_2, and lying, except for the latter endpoint, wholly in D_k. Let S_h and Σ_h, $h = 1, 2, 3, \cdots$, be sequences of subdivisions of K_3 and K_5, respectively, with meshes converging to zero. For a fixed h, only a finite number of the sets D_k can contain vertices of S_h or Σ_h, and consequently we can choose for each h a set $D_{k(h)}$ that contains no vertex of S_h or Σ_h.

For h great enough, all details of the following construction may be carried out. We first consider, progressively, the 1-cells of the subdivision S_h of K_3. Any 1-cell of S_h that lies wholly in D or in M' we replace by the Vietoris 1-chain obtained from its own subdivisions. Suppose ab is a 1-cell of S_h, where a is in D and b is in M'. On ab let a' be the first point of M' in the order from a to b. Then we replace ab by a V-chain based on subdivisions of aa' and a V-chain of M' bounded by a' and b. If a and b both lie in M', we replace ab by a V-chain of M' bounded by a and b. If both a and b lie in D, but ab meets M', we determine a' as above, also a point b' of $M'\cdot ab$ such that $b'b$ has only b' in M', and replace $a'b'$ by a V-chain of M', the portions aa' and $b'b$ of ab being replaced by V-chains based on their own subdivisions. All these operations are carried out progressively, passing from 1-cell to 1-cell, and finally the boundary 1-cycles of S_h are replaced by these 1-chains by the harmonizing process. For h great enough, the 1-chains may be so taken as to form arbitrarily close approximations to the corresponding 1-cells, in the sense that as the diameter of the 1-cells decreases, so too does the maximum diameter of the V-chains that replace them. We note that none of the 1-chains obtained meets $D_{k(h)}$.[21]

We next treat the 2-cells of S_h. Let E^2 be such a 2-cell. Its boundary has been replaced by a V-cycle γ^1 formed of chains bounded by the vertices. Within a small spherical neighborhood of γ^1 let F^2 be a V-chain bounded by γ^1. Denoting F^2 by $\{F_m\}$ and γ^1 by $\{1_m\}$, we have, for each m, $F_m \rightarrow 1_m$. If no F_m meets $D_{k(h)}$, we proceed to another 2-cell. If only a finite number of sets F_m meet $D_{k(h)}$, we may delete those that do. If infinitely many meet $D_{k(h)}$, we may assume that all do, and proceed as follows. By the construction of the preceding paragraph, no point of a cycle 1_m lies in $D_{k(h)}$. Assuming F_m to be a δ_m-chain, which we may do by deleting chain elements from F^2, let G_m denote the sub-

[21] A process similar to that which we use here was also used in the proof of Theorem 8 of G. C. M., except that in the latter case the replacements were carried out in open subsets of E_n and hence ordinary finite chains instead of V-chains could be used.

chain of F_m based on those of its 2-cells that lie in $D_{k(h)}$. Let $E_m = F_m - G_m$. Denoting the boundary of G_m by $1'_m$, we have the relations

$$(6) \qquad \begin{aligned} G_m &\to 1'_m \\ E_m &\to 1_m - 1'_m . \end{aligned}$$

We may assume that the set $\{1'_m\}$ forms a V-cycle, and for h great enough, that it bounds a "small" chain on M'; for by unessential (δ_m)-alterations of $1'_m$, it becomes a cycle of M', and M' as shown above is locally 1-connected. We have, then,

$$(7) \qquad N_m \to 1'_m \qquad \text{on } M'.$$

Combining relations (6) and (7) we get

$$E_m + N_m \to 1_m ,$$

where $E_m + N_m$ fails to meet $D_{k(h)}$.

When we have proceeded through the set of all 2-cells of S_h in this manner, and harmonized so that the chains corresponding to single cell-boundaries form 2-cycles, we are ready to treat the 3-cells of S_h. At this stage none of the 1-chains and 2-chains obtained meets $D_{k(h)}$.

In the final stage of our process of replacement, we utilize the local $(n - 2)$-connectedness of M' to replace the $(n - 1)$-cells of S_h by corresponding $(n - 1)$-chains, which combine to form a cycle Γ_h^{n-1} approximating S_h, the degree of approximation depending only on h (M' being *uniformly* locally i-connected within and on K_1). The cycle Γ_h^{n-1} does not meet $D_{k(h)}$.

Similarly, Σ_h receives, by the same process, an approximating cycle Φ_h^{n-1} which fails to meet $D_{k(h)}$.

As the spheres K_3 and K_5 separate each P_k from the spheres K_1 and K_2, for h great enough the cycle $\Gamma_h^{n-1} + \Phi_h^{n-1}$ separates each P_k from K_1 and K_2. In particular, $P_{k(h)}$ is separated from $Q_{k(h)}$ by $\Gamma_h^{n-1} + \Phi_h^{n-1}$. But the carrier of the latter cycle has no point on $t_{k(h)}$, since otherwise there would be vertices of the cycle in $D_{k(h)}$. Thus the assumption that B is not locally 0-connected leads to a contradiction.

In view of Theorem 7 we might ask whether the boundary of such a domain D as considered therein might not also possess the property of local i-connectedness for $1 \leqq i \leqq n - 2$. That this is not the case may be shown by adding, to the well-known continuum in E_3 consisting of the sphere with infinitely many handles converging to a point, 2-cells in each of its complementary domains which render its 1-dimensional Betti number zero and do not destroy the local 0-connectedness. The resulting continuum is locally i-connected for $i = 0, 1$, and each of its complementary domains has a locally 0-connected boundary, but neither boundary is locally 1-connected. Incidentally, the same example modified so that the handles of the sphere converge to a line segment instead of to a point, shows the necessity for the local 1-connectedness condition of Theorem 7 when $n = 3$.

The result of Theorem 7 suggests considering those compact metric spaces which are locally i-connected for $0 \leqq i \leqq j$ as generalizations of the notion of Jordan continuum, in the sense that although for $j > 0$ the class of spaces so obtained is a subclass of the class of Jordan continua, certain theorems on Jordan continua may find higher dimensional analogues.[22] For this purpose we give the

DEFINITION. A locally compact, metric space M, which, for $0 \leqq i \leqq j$ (j a non-negative integer), is locally i-connected will be called a J^j.

We can then state the following

THEOREM 8. *In E_n, ($n \geqq 2$), let M be a closed point set which is a J^{n-2}. Then if D is a domain complementary to M, the boundary of D is a J^0.*[23]

Proof. As in the proof of Theorem 7, we first show that $E_n - D$ is locally i-connected for $0 \leqq i \leqq n - 2$. Then we suppose that B is not locally 0-connected. It follows that there exist a point P of B and $\epsilon > 0$ such that in $S(P, \epsilon)$ there is a sequence of points $P_1, P_2, P_3, \cdots$ of B having P as sequential limit point and such that no two points P_i, P_j lie in a connected subset of $B \cdot S(P, \epsilon)$. Now let us suppose that $D + B$ is locally 0-connected. Then, since $E_n - D$ is also locally 0-connected, there exists $\delta > 0$ such that any two points of $D + B$ in $S(P, \delta)$ lie in a connected subset of $(D + B) \cdot S(P, \epsilon)$, and any two points of $E_n - D$ in $S(P, \delta)$ lie in a connected subset of $(E_n - D) \cdot S(P, \epsilon)$. Let P_i and P_j be points of the above sequence lying in $S(P, \delta)$. There exists, in $S(P, \epsilon) - B \cdot S(P, \delta)$, a continuum (relative to $S(P, \epsilon)$) K separating P_i and P_j in $S(P, \epsilon)$. Now a connected subset of $D + B$ joining P_i and P_j in $S(P, \epsilon)$ must meet K and hence $K \subset D$. However, there also exists in $S(P, \epsilon)$ a connected subset of $E_n - D$ joining P_i and P_j which must also meet K. As a result of this contradiction we are led to conclude that if B is not locally 0-connected, so, too, is $D + B$ not locally 0-connected.

We can now proceed with the proof as given for Theorem 7, beginning with the definition of the sets H_k of the fourth paragraph.

[22] In the same volume of Fundamenta Mathematicae containing the paper of Borsuk referred to in footnote 2 appears a paper by C. Kuratowski (pp. 269–287) suggesting this same sort of generalization, using, however, the type of local connectedness defined in Lefschetz's *Topology*. It would be interesting to know if certain of Kuratowski's results, particularly such as are obtained in section II, are obtainable when local connectedness in the sense of V-cycles is assumed.

[23] It will be noted that we do not need to assume either compactness or connectedness for M.

UNIVERSITY OF MICHIGAN.